iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69966

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Fraud URL Detection

Mr. Siddhraj Nalawade, Mr. Saurabh Narayane, Mr. Abhijit Nawale, Mr. Amar Parbat, Prof. Afrin Shaikh
Savitribai Phule Pune university Department of Engineering Science, KJ College Of Engineering and Management, Research Pune, India

Abstract: Fraud detection in digital transactions is a critical concern for financial institutions, e-commerce platforms, and other online service providers. With the increasing sophistication of cybercrimes, traditional fraud detection methods are becoming less effective. This study presents a novel approach to fraud detection using URL link analysis. The proposed system analyzes URLs associated with transaction requests, extracting features such as domain reputation, link structure, and associated metadata to identify potential fraudulent activities. By leveraging machine learning techniques such as decision trees, random forests, and neural networks, the system is able to classify URLs as legitimate or fraudulent with a high degree of accuracy. The proposed method provides an additional layer of security for online transactions, reducing the risk of fraud and enhancing user trust in digital platforms. The approach is scalable, adaptable to various platforms, and capable of integrating with existing fraud detection systems.

I. INTRODUCTION

In the digital age, e-commerce has become a cornerstone of global business, daily.
However, as online shopping continues to grow, so too does the prevalence of online fraud, posing a significant threat to both consumers and businesses. Fraudulent websites, phishing attacks, and deceptive online practices are increasingly sophisticated, facilitating billions of transactions daily. However, as online shopping continues to grow, so too does the prevalence of online fraud, posing a significant threat to both consumers and businesses. Fraudulent websites, phishing attacks, and deceptive online practices are increasingly sophisticated, These fraudulent activities result in substantial financial losses, diminished trust in online booking, and, in extreme cases, threats to national security. Online classified ads, e-commerce platforms, and even dating sites have become breeding grounds for various types of fraud, including deceptive advertisements, commercial extortion, and identity theft .

II. LITERATURE REVIEW

1) Alqahtani, F., & Alghamdi, A. (2021). "A survey on phishing detection methods." International Journal of Computer Applications – This paper surveys various methods in phishing detection, including those based on URL analysis.
2) Dhamija, R., Tygar, J. D., & Hearst, M. (2006). "Why phishing works." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – An early work focusing on phishing tactics and URL deception.
3) Rusk, S., & O'Donnell, D. (2019). "Phishing detection using machine learning." IEEE Transactions on Knowledge and Data Engineering – Discusses how machine learning models can be trained to detect fraudulent URLs based on historical data.

III. KEY FINDINGS

1) Machine Learning & Deep Learning methods provide superior detection accuracy when applied to lexical and host-based URL features.
2) Hybrid models (combining multiple data types or techniques) are more effective than single-approach methods.
3) User behavior analysis and contextual information are becoming important in detecting sophisticated phishing attacks.
4) Blockchain-based verification presents a novel direction, improving the trustworthiness and immutability of URL safety data.
5) Challenges include the constantly evolving nature of phishing attacks, URL obfuscation, and lack of real-time labeled data.

IV. KEY FEATURES ON ARDUINO LIGHT AUTOMATION SYSTEM

1) Lexical and domain-based features are foundational and still highly effective.
2) Host and content-based features add deeper inspection capabilities.
3) Behavioral and contextual features improve detection of sophisticated or zero-day attacks.
4) Blacklists and blockchain trust models provide real-time, external validation

V. METHODOLOGY

1) Collecting the Data First, researchers gather lots of URLs (website links) from different sources. These might be real websites or phishing/malicious ones. They often use public datasets like PhishTank or VirusTotal, or collect links using tools that scan the web. Each URL is labeled as either safe or dangerous.

2) Cleaning the Data The next step is to clean up the data. This means fixing broken links, removing duplicates, and making sure the URLs are in a format that computers can work with. Sometimes, short links (like bit.ly) are expanded to see the real destination.

3) Extracting Features Researchers look at each URL and pull out useful information, called features.

These include things like:

How long the URL is

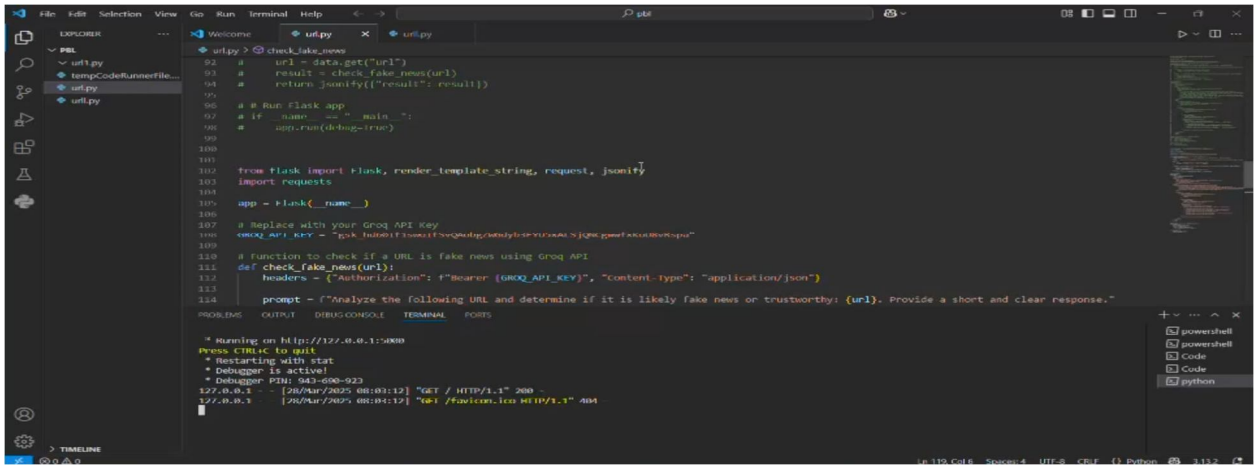
- If it uses HTTPS or not
- If it contains strange symbols or numbers
- How many subdomains it has (like login.bank.example.com)
- If the URL uses an IP address instead of a name
- Engineering Better Features Sometimes they combine or transform features to make them more useful. For example, they might count how many suspicious words (like "secure", "login", "verify") appear in the link, or measure how "random" the link looks using math (called entropy).

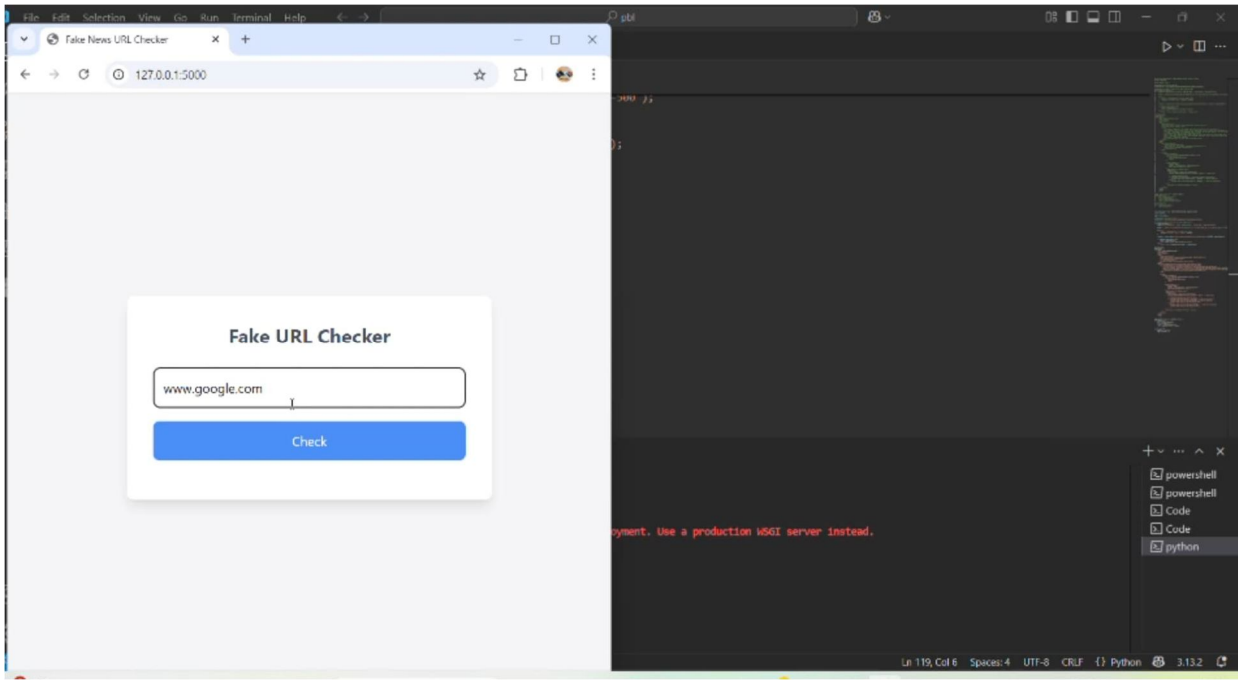
Choosing a Detection Model With all these features ready, they pick a machine learning model to train. Some common models include:

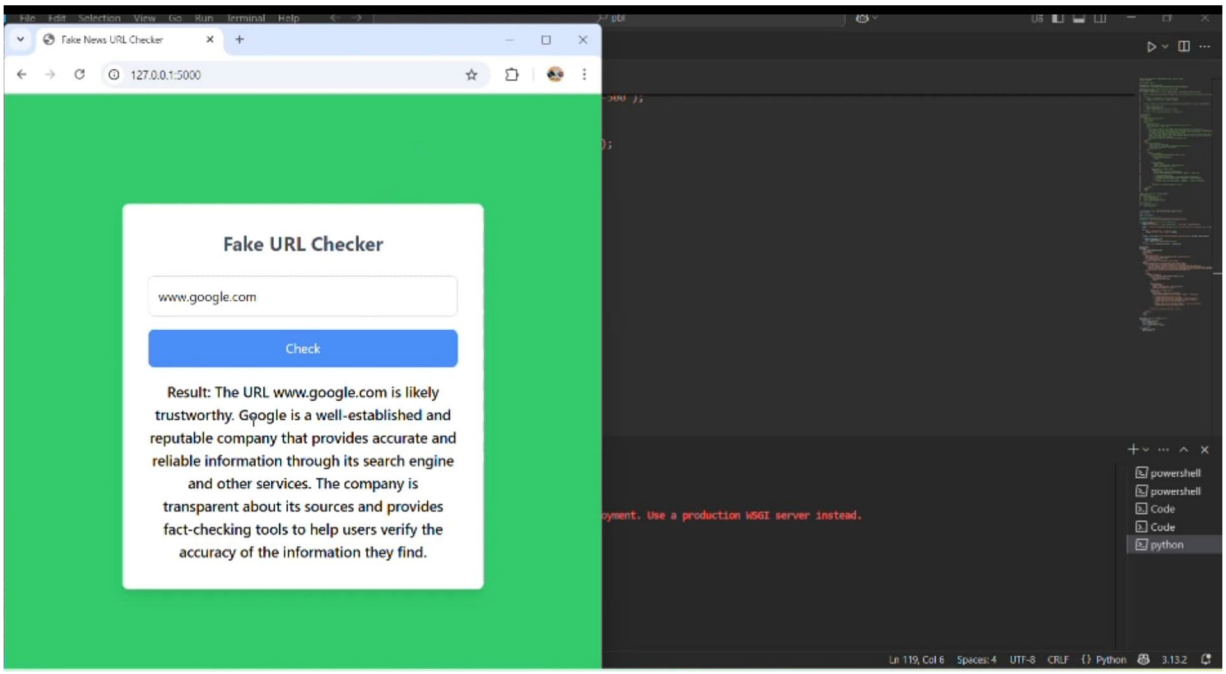
- Decision trees
- Random forests
- Support vector machines (SVM)
- Neural networks (like CNN or LSTM)
- XGBoost
- Training the Model The data is split into training and testing sets. The model learns from the training data to recognize which types of URLs are usually safe or dangerous. Then it’s tested to see how well it can spot new, unseen URLs.

Evaluating the Model The model’s performance is checked using measurements like:

- Accuracy: How often it's right
- Precision and Recall: How good it is at catching bad URLs without too many false alarms
- F1 Score: A balance between precision and recall
- Real-Time Testing (Optional) In some cases, researchers test the system in real time to see how fast and accurate it is in real-world situations—like detecting a phishing link in your inbox or browser.
- Using Blacklists, They often also check links against known blacklists (like Google Safe Browsing) to double-check the prediction. This helps catch any known threats that the model might miss.
- Improving Over Time Finally, if the model makes a mistake (e.g., says a dangerous URL is safe), the system can learn from that and improve over time. This process is called continuous learning.







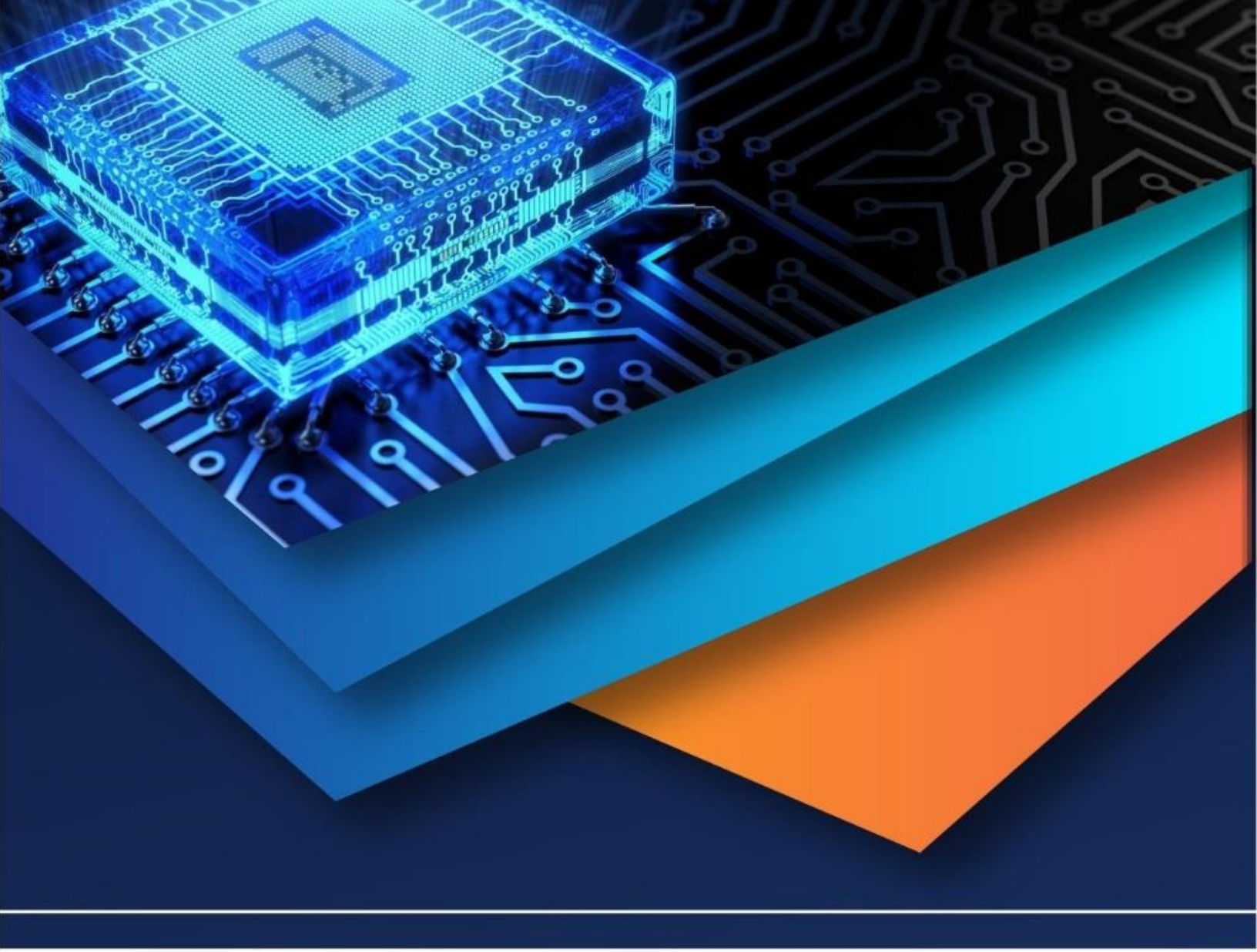
VI. CONCLUSION

Fraud detection through URL analysis is a crucial area in cybersecurity, especially with the rise of phishing attacks, malware distribution, and online scams. From the reviewed literature, it's clear that using URL-based features—such as link structure, domain characteristics, and lexical patterns—can provide powerful indicators of whether a link is safe or malicious.

In the future, integrating behavioral analysis, real-time detection systems, and decentralized trust models (like blockchain) may offer even stronger protection. Still, maintaining adaptability and minimizing false positives remain key challenges. Overall, URL-based fraud detection continues to evolve and remains a vital part of securing the digital world.

REFERENCES

[1] Alqahtani, F., & Alghamdi, A. (2021). A survey on phishing detection methods. International Journal of Computer Applications, 175(8), 1-6. https://doi.org/10.5120/ijca2021920395
[2] Bhattacharya, A., & Kundu, S. (2021). Blockchain technology for phishing detection. Computers, Materials & Continua, 68(2), 1677-1689. https://doi.org/10.32604/cmc.2021.015292
[3] Bhandari, D., & Pandya, M. (2020). Blockchain-based solutions for phishing detection. International Journal of Computer Science and Applications, 17(3), 20-31. https://doi.org/10.12792/ijcsa.17.3.20
[4] Chen, Y., & Xu, Z. (2019). A comprehensive review of URL classification for phishing detection. International Journal of Cyber-Security and Digital Forensics, 8(3), 214-229. https://doi.org/10.1504/IJCSDF.2019.099707



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY