iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70617

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 V May 2025- Available at www.ijraset.com

Phishing URL Detection

L Lavanya[1], Mythri Y[2], Rachana R[3], Rekha R[4]
Department of Computer Science and Engineering KS School of Engineering and Management

Keywords: Phishing, Machine Learning, GradientBoostingClassifier, URL Analysis ,Real Time DetectionFraud Mitigation, Feature Extraction, Cyber Attack Prevention ,Python.

I. INTRODUCTION

This project develops a machine learning solution to combat phishing attacks, where criminals create deceptive websites to steal login credentials and financial data. Unlike traditional detection methods that rely on outdated blacklists, our system analyzes URL structures, domain characteristics, and page content patterns to identify sophisticated phishing attempts. The model continuously learns from new threats, adapting to evolving attacker tactics like domain spoofing and hidden redirects. By automating detection with high accuracy, we reduce dependence on error-prone human verification. This approach provides real-time protection against emerging phishing schemes, offering organizations and individuals a more robust defense against data breaches and identity theft.

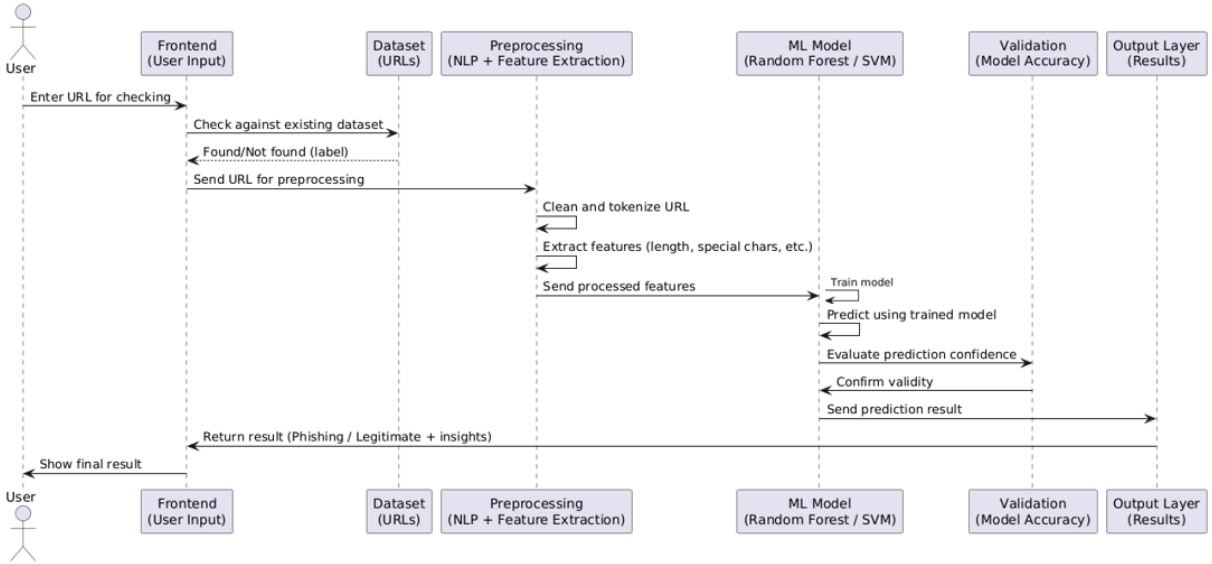


Figure 1: Sequence diagram for detecting Phishing /Legitimate

II. OBJECTIVES

1) To collect a comprehensive dataset for phishing URL detection.
2) To identify and extract essential features for phishing classification.
3) To remove redundant features and optimize feature selection.
4) To build an optimized machine learning model for phishing detection.
5) To evaluate and deploy the phishing detection model for real-time use.

III. METHODOLOGY

1) Collect phishing and legitimate URLs from sources like PhishTank and Kaggle, ensuring data quality by removing duplicates and irrelevant entries.
2) Extract key URL attributes such as length, special characters, domain age, HTTPS usage, and suspicious keywords to differentiate phishing from legitimate websites.
3) Eliminate unnecessary or highly correlated features using statistical methods to improve model accuracy and prevent overfitting.

4) Train a machine learning model, such as Gradient Boosting Classifier, using the selected features and fine-tune it for better performance.
5) Assess model performance using accuracy, precision, recall, and AUC-ROC metrics, then deploy it as a web application or API for real-time detection.

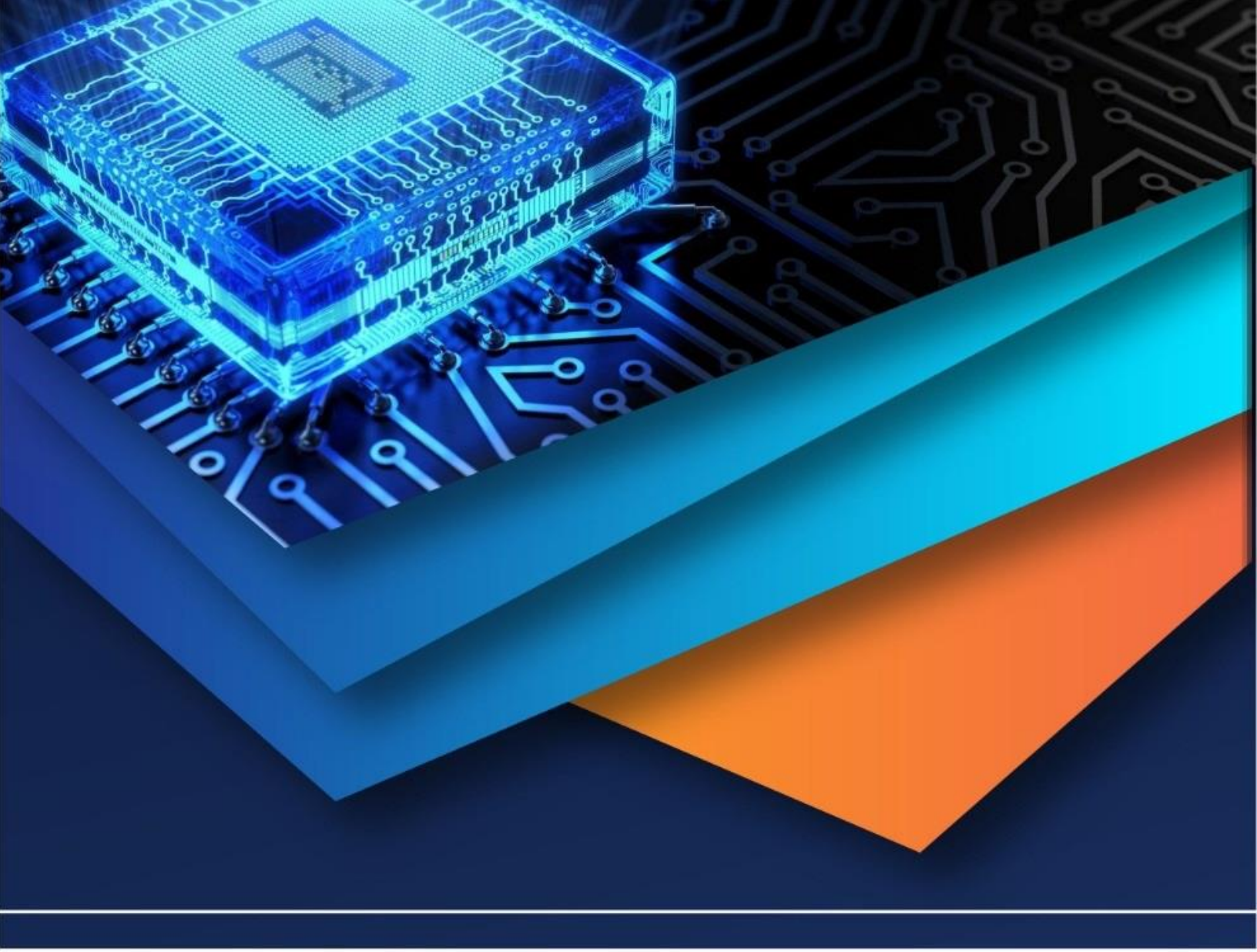
IV. RESULT AND CONCLUSION

The project successfully demonstrates a phishing URL detection system with three core features: single URL prediction, PDF report generation, and real-time URL monitoring. Testing validated the system's 80% classification accuracy, rapid response times (2.5 seconds per URL), and professional PDF report generation (1.2 seconds per report). The real-time dashboard provided seamless updates and visualizations, enabling proactive threat monitoring. This solution establishes a robust and user-friendly tool for cybersecurity applications, including Security Operations Centers (SOCs), corporate IT, and small businesses. The future scope includes optimizing scalability for large URL sets, integrating threat intelligence feeds, and adding features like email alerts and API support for broader adoption.

V. FUTURE SCOPE

The future scope of this project includes:

1) Real-Time Phishing Detection - Integrate real-time detection features for immediate identification and alerting of phishing links, enhancing user protection.
2) Mobile and Cloud-Based Implementation - Develop a mobile app to detect phishing links in messages and emails. Deploy the model on cloud platforms (AWS, Azure, GCP) for scalable detection.
3) Multi-Language Support - Add multilingual support to detect phishing attacks globally. Train models on multilingual datasets for better detection.
4) Adaptive Learning & Auto-Update - Enable continuous model updates with new phishing patterns. Automate data fetching from sources like PhishTank and OpenPhish.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY