Project Report

Project Report: Phishing Website Detection System

1. Introduction:

Phishing attacks are a growing concern in the digital world. This project focuses on developing a real-time phishing website detection system that can classify URLs as either Legitimate or Phishing using machine learning techniques.

2. Objectives:

- To detect phishing websites using URL-based features.

- To build a machine learning model capable of real-time classification.

- To deploy the solution as an accessible web application.

3. Dataset and Feature Engineering:

A dataset of URLs labeled as either 'phishing' or 'legitimate' was used. 20 URL-based features were engineered such as:

- Presence of '@' symbol

- Length of URL

- Use of HTTPS

- Number of subdomains

- Presence of IP address instead of domain name

4. Preprocessing:

Custom min-max normalization was implemented without external libraries to ensure deployment simplicity and efficiency.

5. Model Training:

A Random Forest Classifier was chosen due to its robustness and accuracy. The model was trained and validated, achieving high accuracy on unseen data.

6. Application Development:

The backend was developed using Flask. The web application allows users to input URLs, which are then processed and classified in real-time.

7. Deployment:

The complete application was deployed using Render, allowing public access and real-time phishing detection.

8. Results:

- High accuracy and precision achieved using Random Forest.

- Real-time classification capability tested successfully.

9. Learning Outcomes:

- Hands-on experience in machine learning, web development, and deployment.

- Developed an end-to-end understanding of applied AI systems.

10. Future Scope:

- Integration with browser extensions for direct URL scanning.

- Adding NLP-based content analysis for higher accuracy.

- Continuous retraining with fresh data to maintain performance.