Phishing Domain Detection System

1. Introduction

This document outlines the high-level design for a Phishing Domain Detection System.

The system aims to leverage machine learning to distinguish between real and malicious domains.

2. System Requirements

Functional Requirements:

- Predict whether a given domain is real or fake using Scikit-learn.

- Allow for model testing through an API or user interface.

Non-Functional Requirements:

- Code: Modular, Safe, Testable, Maintainable, Portable

3. System Architecture

- Data Acquisition: Downloads and preprocesses the phishing domain dataset.

- Data Preprocessing: Cleans and prepares data.

- Feature Engineering: Extracts features from URLs and domains.

  - URL-Based Features (length, special characters)

  - Domain-Based Features (age, registration info)

- Model Building: Trains ML models (e.g., Random Forest, SVM)

- Model Evaluation: Uses accuracy, precision, recall

- Model Selection: Deploys best model

- API/UI: Allows user testing of the model

4. Technology Stack

Phishing Domain Detection System

- Python

- Scikit-learn

- Flask / Flask-RESTful

5. Data Flow

1. Data downloaded

2. Preprocessed

3. Features extracted

4. Models trained

5. Evaluated

6. Best model selected

7. User inputs URL

8. Prediction shown

6. Success Criteria

- Accurate classification

- Modular, maintainable system

- Easy-to-use interface

7. Conclusion

This system helps detect phishing domains using machine learning, improving digital safety.