High-Level Design (HLD) Document

Phishing Detection System

Table of Contents

---

1. Introduction

The Phishing Detection system aims to predict whether a URL is a Phishing URL or not. This High-Level Design (HLD) document provides an overview of the system architecture and the major components involved in the project.

2. System Architecture

The system architecture for this project consists of the following components:

1. User Interface
2. Application Server
3. Machine Learning Model
4. Database

3. User Interface

The User Interface component provides a user-friendly interface for users to interact with the system. It allows users to input required information in the form. The User Interface component sends the input data to the Application Server for further processing.

4. Application Server

The Application Server acts as the central processing unit of the system. It performs the following tasks:

- **Data Ingestion:** The Application Server receives the input data from the User Interface and validates its integrity and quality. It ensures that the required fields are present and handles any missing or erroneous values.
- **Data Transformation:** The Application Server preprocesses the input data by performing feature engineering, data normalization, and encoding categorical variables. It prepares the data for input into the Machine Learning Model.
- **Model Prediction:** The Application Server utilizes the Machine Learning Model to make predictions based on the preprocessed input data. It sends the input data to the Model and receives the predictions as the output.
- **Data Persistence:** The Application Server stores the input data and corresponding predictions in the Database for future reference and analysis.
- **UI Application:** The Application Server receives the predicted data and shows to the user on the UI.

5. Machine Learning Model

Model is trained using Phishing URL dataset from kaggle. The Model leverages various machine learning algorithms, such as Random Forest, SVM, Gradient Boosting, to learn patterns and make accurate predictions. The Model is trained and periodically updated to adapt to changing data patterns and improve its performance.

6. Database

The Database component stores the input data and the corresponding predictions made by the Application Server. It provides data persistence and allows for data retrieval and analysis. This project uses MongoDB to ensure data integrity and accessibility.

7. Documentation and Collaboration

Comprehensive documentation, including code documentation, system architecture, data dictionaries, and user manuals, is maintained to facilitate system understanding, maintenance, and future enhancements.

Code on GitHub

8. Conclusion

The High-Level Design (HLD) document provides an overview of the system architecture and the major components involved in the Phishing Detection System in the e-commerce domain. It outlines the interaction between the User Interface, Application Server, Machine Learning Model, and Database components. The document serves as a guide for the development, implementation, and maintenance of the system, ensuring a structured and coherent approach.