Phishing Website Detection System - High Level Document

Project Title:

Phishing Website Detection System

Internship Organization:

Physics Wallah

Role:

Machine Learning Intern

Duration:

[Insert Duration, e.g., May 2025 - June 2025]

Project Overview:

This project involves the design and development of a web-based system that can accurately classify a given URL as Phishing or Legitimate. It aims to provide users with a fast, reliable tool to detect fraudulent websites and prevent cyber-attacks such as identity theft, financial fraud, and credential leaks.

Objectives:

- Build a machine learning model to classify URLs
- Develop a lightweight feature extraction engine
- Normalize data using a custom min-max approach
- Create a web interface using Flask
- Deploy the solution on Render

Tech Stack:

Programming Language: Python

ML Library: Scikit-learn

Phishing Website Detection System - High Level Document

Web Framework: Flask

Deployment: Render

Model Used: Random Forest Classifier

Feature Extraction: Custom URL parsing and domain inspection logic

Feature Engineering:

The system uses 20 top-performing features selected based on their contribution to model performance. Key features include:

- directory_length, time_domain_activation, file_length, ttl_hostname, asn_ip, time_response, etc.

Min-max normalization was applied using custom logic to ensure feature values stay within [0, 1].

Model Workflow:

1. User inputs URL via the web interface
2. Features are extracted in real-time
3. Features are manually normalized
4. Pre-trained Random Forest model classifies the input
5. Result is displayed instantly

Deployment:

- Web application deployed on Render
- Model and logic bundled into a .joblib file
- Flask-powered URL prediction endpoint

Impact:

- Helps users avoid phishing attacks
- Promotes cybersecurity awareness
- Demonstrates real-world ML security application

Files Included:

Phishing Website Detection System - High Level Document

- model_training.ipynb
- extract_features.py
- predict.py
- app.py
- index.html
- scaled_model.joblib