Project design phase II

Data flow diagram

Date	28 September 2022
Team ID	PNT2022TMID07498
Project name	Web phishing detection
Maximum marks	4 marks

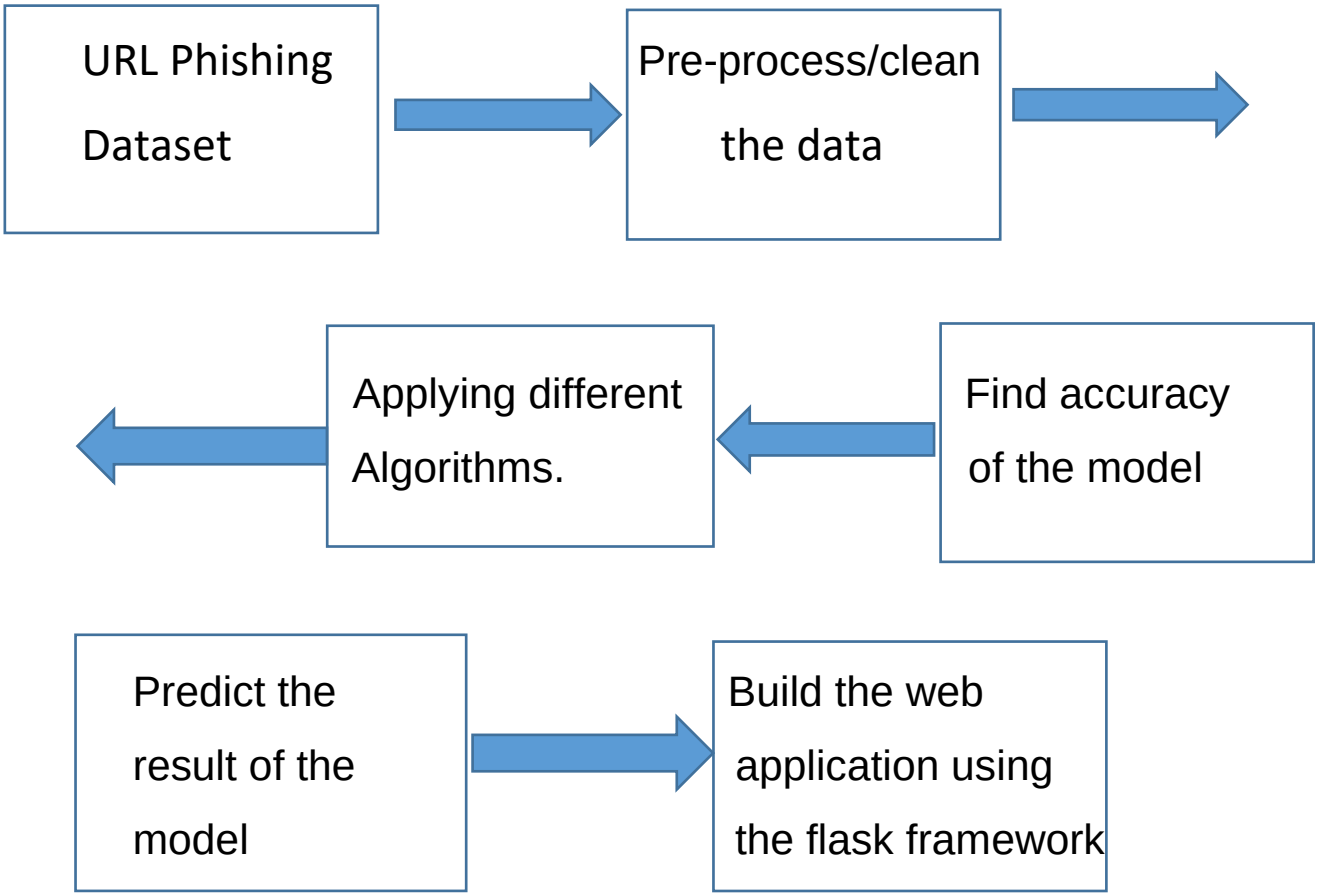
WEB PHISHING DETECTION:

Phishing URL may be a widely used and customary technique for cybersecurity attacks. Phishing is a cybercrime that tries to trick the targeted users to expose their private and sensitive information to the attacker. The motive of the attacker is to gain access to personal information like usernames, login credentials, passwords, financial account details, social networking data, and personal addresses.

These private credentials are then often used for malicious activities like fraud, notoriety, gain, reputation damage, and lots of more illegal activities. This paper presents a comprehensive study of various existing systems used for phishing website detection. The system presented here uses advanced machine learning and to realize better precision and better accuracy while categorizing websites as phishing or begin.

With the booming of internet transactions, cybercrimes rapidly increased and with anonymity presented by the internet, Hackers attempt to trap the end-users through various forms such as phishing, SQL injection, malware, man-in-the-middle, domain name system tunnelling, ransomware, web trojan, and so on. Among all these attacks, phishing reports to be the most deceiving attack Typically, the intent is to get users to reveal financial information, system credentials or other sensitive data.

DATA FLOW DIAGRAM OF WEB PHISHING DETECTION



URL Phishing dataset:

The dataset consists of a collection of legitimate as well as phishing website instances. Each instance contains the URL and the relevant HTML page. The index. sql file is the root file, and it can be used to map the URLs with the relevant HTML pages. The dataset can serve as an input for the machine learning process.

Pre-process/clean the data:

Data cleaning is the process of adding missing data and correcting, repairing, or removing incorrect or irrelevant data from a data set. Dating cleaning is the most important step of pre-processing because it will ensure that your data is ready to go for your downstream needs

Applying different algorithms:

Algorithms are normally built-in underlying languages, that means it can be carried out in more than one programming language. Algorithms are used as specifications for data processing, doing mathematics, automated reasoning, and several other chores like this.

Find the accuracy of the model:

Accuracy is a metric used in classification problems used to tell the percentage of accurate predictions. We calculate it by dividing the number of correct predictions by the total number of predictions.

Build web applications using the Flask framework:

- Installing Flask.
- Creating a Base Application.
- Using HTML templates.
- Setting up the Database.
- Displaying All Posts.