Ideation Phase

Literature Survey

Date	19 September 2022
Team ID	PNT2022TMID24978
Project Name	Project – Web Phishing
Maximum Marks	2 Marks

Web Phishing Detection

Abstract

It is a crime to practice phishing by employing technical tricks and social engineering to exploit the innocence of unaware users. Phishing is a common attack against Internet users that causes them to reveal their information using fake websites. The goal of the fake website is to steal personal information such as usernames, passwords and online banking transactions. Scammers use websites that are visually and semantically similar to the real ones. As technology continues to advance, phishing techniques begin to advance rapidly, and this should be prevented by using anti-phishing mechanisms such as spoofed URL detection. Machine Learning is a powerful tool used to combat spoofing attacks. This report covers machine learning technology to detect fake URLs by extracting and analysing different characteristics of legitimate and fake URLs. Random Forest, Logistic Regression and algorithms are used to detect fake websites.

Introduction

The purpose or goal behind phishing is data, money or personal information stealing through the fake website. The best strategy for avoiding the contact with the phishing web site is to detect real time malicious URL. Phishing websites can be determined on the basis of their domains. They usually are related to URL which needs to be registered (low-level domain and upper-level domain, path, query). Recently acquired status of intra-URL relationship is used to evaluate it using distinctive properties extracted from

words that compose a URL based on query data from various search engines such as Google and Yahoo**.** These properties are further led to the machine-learning based classification for the identification of phishing URLs from a real dataset. It can also be defined as intentionally using harsh weapons such as Spasm to automatically target the victims and targeting their private information.

Literature Survey

1. Protecting user against phishing using Anti-phishing: -

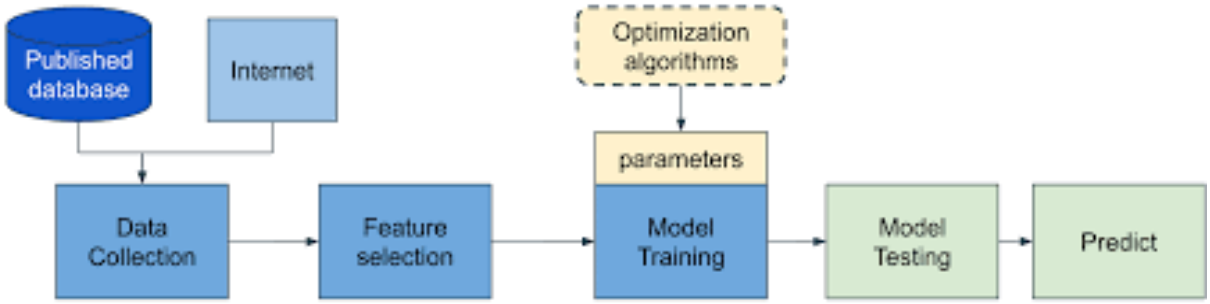
Anti-Phish is used to avoid users from using fraudulent web sites which in turn may lead to phishing attack. Here, Anti-Phish traces the sensitive information to be filled by the user and alerts the user whenever he/she is attempting to share his/her information to a untrusted web site. The much effective elucidation for this is cultivating the users to approach only for trusted websites.

2. Learning to Detect Phishing Emails:

An alternative for detecting these attacks is a relevant process of reliability of machine on a trait intended for the reflection of the besieged deception of user by means of electronic communication. This approach can be used in the detection of phishing websites, or the text messages sent through emails that are used for trapping the victims.

3. Phishing detection system for e-banking using fuzzy data mining: -

Phishing websites, mainly used for e-banking services, are very complex and dynamic to be identified and classified. Due to the involvement of various ambiguities in the detection, certain crucial data mining techniques may prove an effective means in keeping the e-commerce websites safe since it deals with considering various quality factors rather than exact values.



FEATURES OF PROPOSED SYSTEM:

1. **FUNCTIONAL CAPABILITIES:**

   The ultimate aim of this project is to detect phishing attacks in real-time. This model checks the website with machine learning server for any maliciousness in the accessed site.

2. **PERFORMANCE LEVEL:**

   At the client side, it takes 1-2 seconds to detect whether a site is phishing or not.

3. **DATA STRUCTURES:**

   The data in this project are maintained in the CSV form. It provides easy access to the user.

4. **SAFETY:**

   No data loss occurs in this system.

5. **RELIABILITY:**

   We assure that the project is completely authenticated in order to enhance security and corruptions of database as well as the software.

6. **QUALITY:**

   The project is developed with the help of Anaconda Navigator software which meets the requirement of the user, the project is checked whether the phases individually have a served its purpose.