Web Phishing Detection Using Ibm Watson

SmartInternz Guided Project

Report

Submitted by :

1. Anamika Lochab - 18BCE10035
2. Arpita Pandey - 18BCE10060
3. Vitti Gupta - 18BCE10299
4. Vishwas - 18BCG10104
5. Sneha Rani - 18BOE10060

1. Introduction

a. Overview

Phishing is a digital assault which assaults the client's close to home data like account id, email subtleties, any close to home passwords and so on. The assailants fool the clients like they accept that the connection is reliable and we can fill the subtleties of our ledger or anything. There are numerous enemies of phishing arrangements which incorporate boycott or white list, heuristic and noticeable closeness-based systems proposed to date, however online clients are all things considered getting caught into uncovering touchy insights in phishing sites. A principle novel characterization is mostly founded on the heuristic highlights that are created from the URL, source code and hardly any outsider administrations to redress the issues of the prior phishing systems. The model that has been proposed now is an intelligent, flexible and effective system that is based on using classification algorithms. The implemented classification algorithms and techniques do extract the phishing datasets criteria to classify their legitimacy.

b. Purpose

This project aims to present a framework to detect phishing websites using Machine Learning Approach. This method can detect up to 91.67% of phishing websites. Phishing is a type of fraud to access users' credentials. The attackers access users' personal and sensitive information for monetary purposes. Phishing affects diverse fields, such as e-commerce, online business, banking and digital marketing, and is ordinarily carried out by sending spam emails and developing identical websites resembling the original websites. As people surf the targeted website, the phishers hijack their personal information.

2. Literature Survey

a. Existing problem

Phishing is a growing problem nowadays which requires greater defence. Phishing is the most popular attack vector for criminals and has grown 65% in the last year, according to Retruster. The problem with phishing is that attackers constantly look for new and creative ways to fool users into believing their actions involve a legitimate website or email. Phishers have become more skilled at forging websites to appear identical to the expected location, even including logos and graphics in the phishing emails to make them more convincing. Malicious links will lead to a website that often steals login credentials or financial information like credit card numbers. Attachments from phishing emails can contain malware that once opened can leave the door open to the attacker to perform malicious behaviour from the user's computer.

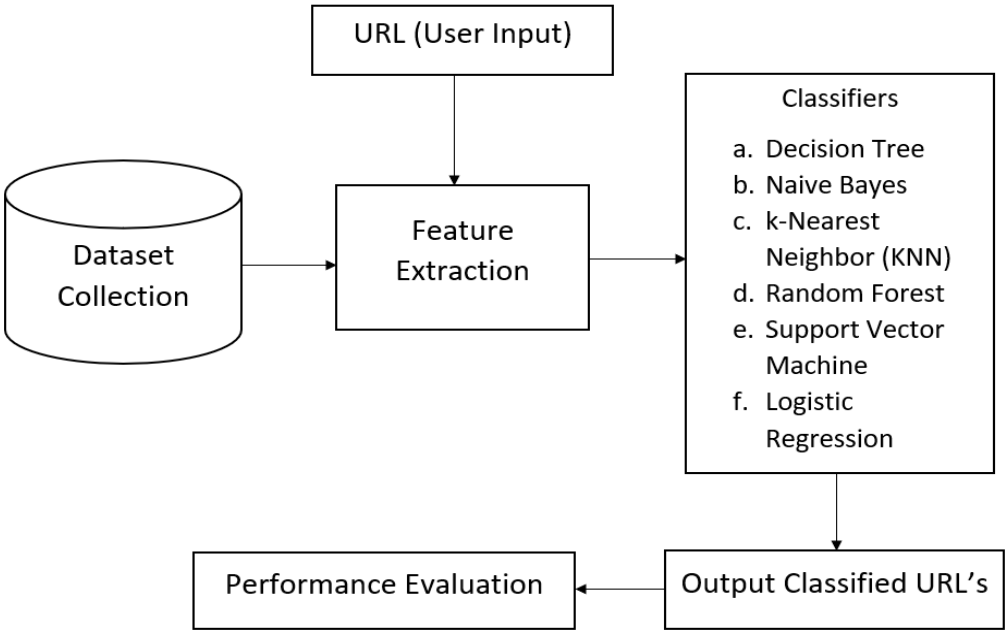
b. Proposed solution

We propose that the possible solution to web phishing is to detect phishing attacks before they reach the user .The URL of phishing websites may be very similar to real websites to the human eye, but they are different in IP. Our solution uses different machine learning models trained over features like if URL contains @, if it has double slash redirecting, page rank of the URL, number of external links embedded on the webpage, etc. This approach could get upto 92% accuracy for detecting whether the site is legitimate or not.

3. Theoretical Analysis

a. Block diagram

The following is the block diagram of our proposed solution to predict whether the website is legitimate or not.



b. Hardware / Software designing

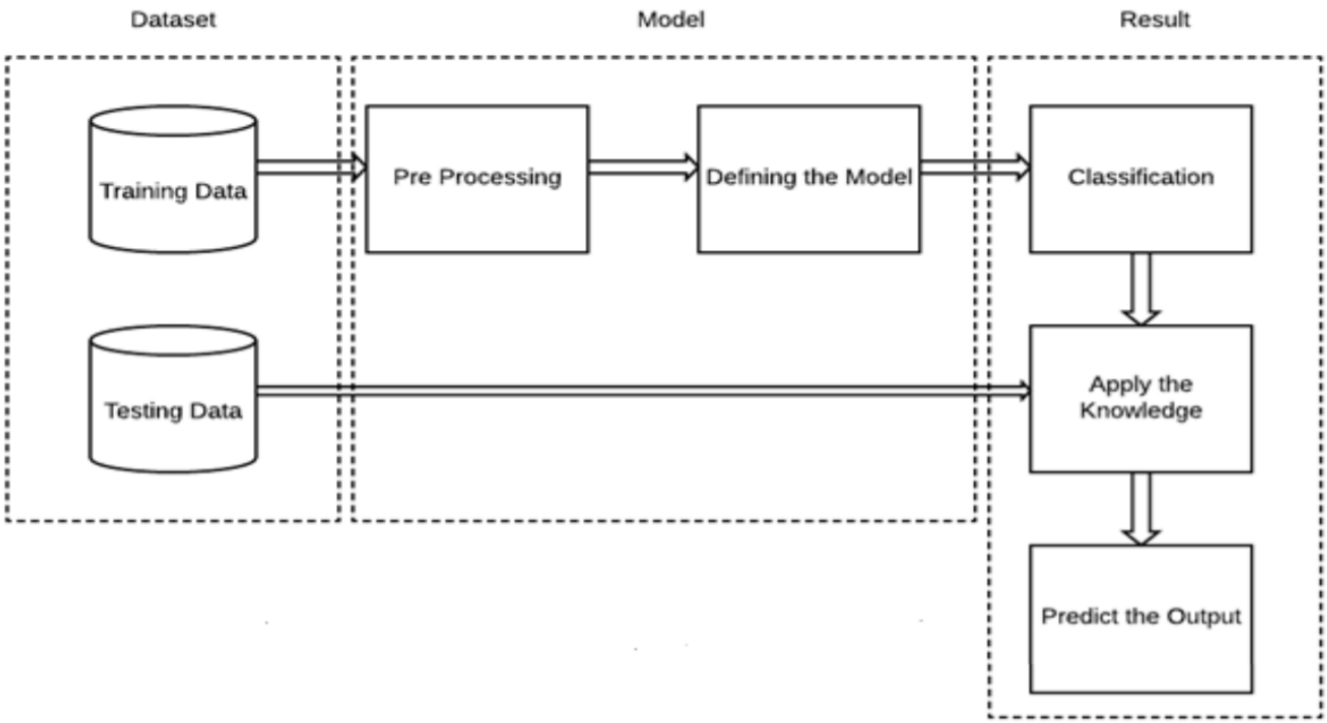
Hardware requirements : Laptop

Software requirements : Python - 3.6, Spyder, Jupyter Notebook

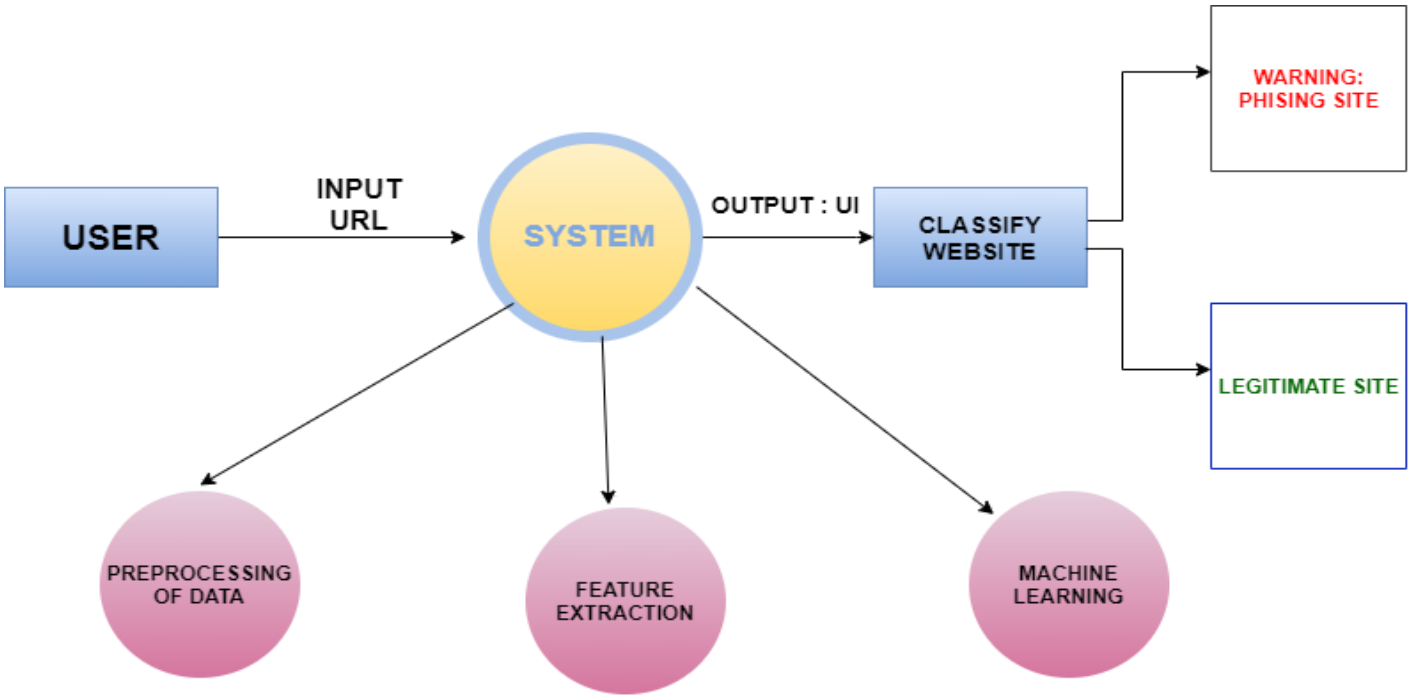
4. Experimental Investigations

- It showed that there is a significant relation between the two phishing website criteria and for identifying phishing websites.
- The experiments show that the classifiers were successful in distinguishing real websites from fake ones over 90% of the time.

5.Flowchart



6.Result



In this project, we have successfully developed an intelligent, flexible, and effective system based on classification algorithms for detecting and

predicting e-banking phishing websites.The Web application was built using a flask that integrates with the model built. In the application, the user provides any website URL to check and the corresponding parameter values are generated by analysing the URL using which legitimate websites are detected.

7.

Advantages

- Use of new classification features and algorithms with improved accuracy
- More adaptable
- Increased accuracy and decreased false positive rate
- Provide secure and healthy online shopping and e-banking environment to the users
- Do not require changes in authentication platforms
- Do not rely on the user's ability to detect phishing
- Easy deployment of our phishing detection model to end users

Disadvantages

- Require frequent training over new features with time
- Needs computational resources and time for training
- Cannot be solely implemented at client side due to resource limitations
  over browsers
- High Storage complexity as there is need to store and update training dataset
- Frequent need for training of new features if phishers start bypassing them

8. Applications

- Detect and predict e-banking phishing websites, thus allowing safe online purchasing and payment through e-banking
- The output provided as a user friendly web platform which can further be extended to a browser extension
- Easy deployment of our phishing detection model to end users
- To successfully identify the phished website looking similar to the original website as malicious and thus prevent phishing attacks vulnerable to users in the online space
- Prevent other types of web phishing attacks, malicious links and avoid deception
- Minimal user training required and does not require any changes to the existing authentication schemes

9. Conclusion

In this project, we built a mechanism to detect phishing websites. Our methodology uses not just traditional URL based or content based rules but rather employs the machine learning technique to identify not so obvious patterns and relations in the data. We have used features from various domains spanning from URL to HTML tags of the webpage, from embedded URLs to favicon. To check the traffic and status of the website. We were able to obtain an accuracy of more than 91% thus classifying most websites correctly and proving the effectiveness of the machine learning we are using Logistic Regression technique to attack the problem of phishing websites. We provided the output as a user-friendly web platform which can further be extended to a browser extension to provide safe and healthy online space to the users.

10. Future Scope

The platform can be converted into a browser extension. Phishing websites usually inflict losses to the users by acting as clickbaits. So, a browser extension can help prevent accidental land ups on these websites, by checking every URL which the browser tries to open, before actually allowing the user to land up on the page.

11. Bibliography

[1] Ankit Jain and B B Gupta. Phishing detection: Analysis of visual similarity based approaches. Security and Communication Networks, 2017:1–20, 01 2017.

[2] A. Alswailem, B. Alabdullah, N. Alrumayh, and A. Alsedrani. Detecting phishing websites using machine learning. In 2019 2nd International Conference on Computer Applications Information Security (ICCAIS), pages 1–6, 2019.

[3] R. M. Mohammad, F. Thabtah, and L. McCluskey. An assessment of features related to phishing websites using an automated technique. In 2012 International Conference for Internet Technology and Secured Transactions, pages 492–497, 2012.

[4] R. M. Mohammad, F. Thabtah, and L. McCluskey. UCI machine learning repository, 2012

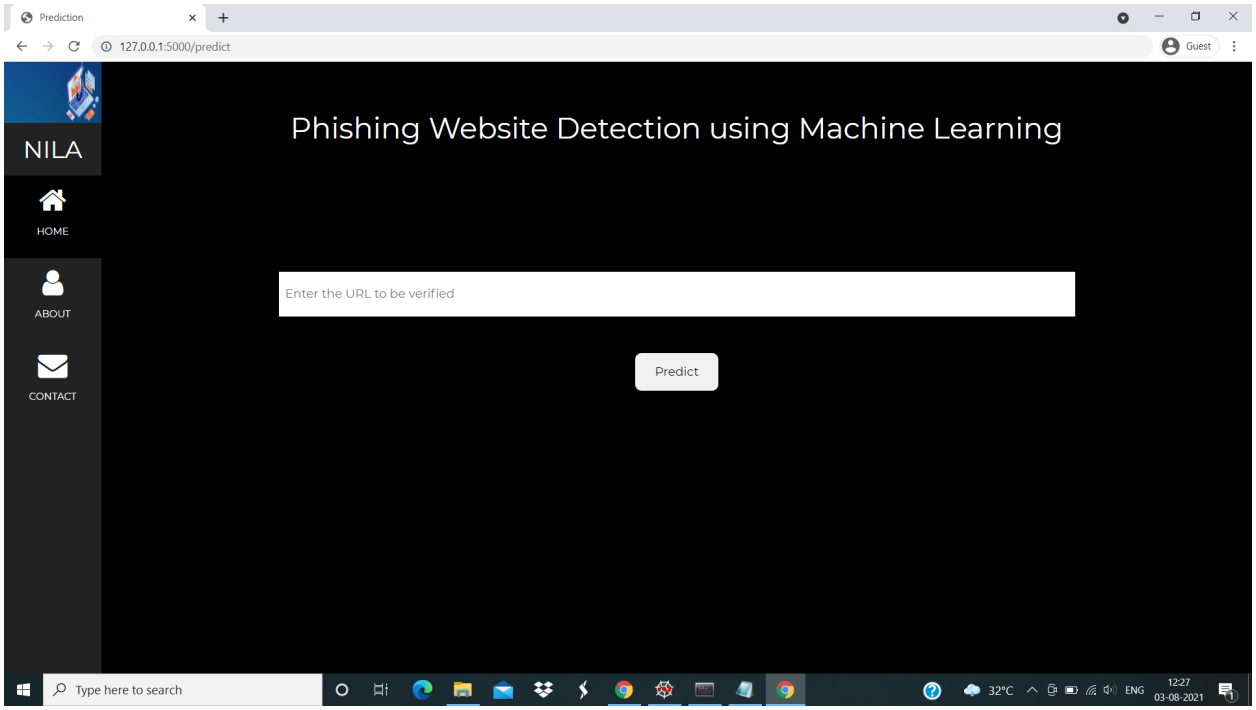
[5] M E Pratiwi, T A Lorosae, and F W Wibowo. Phishing site detection analysis using artificial neural networks. Journal of Physics: Conference Series, 1140:012048, dec 2018.

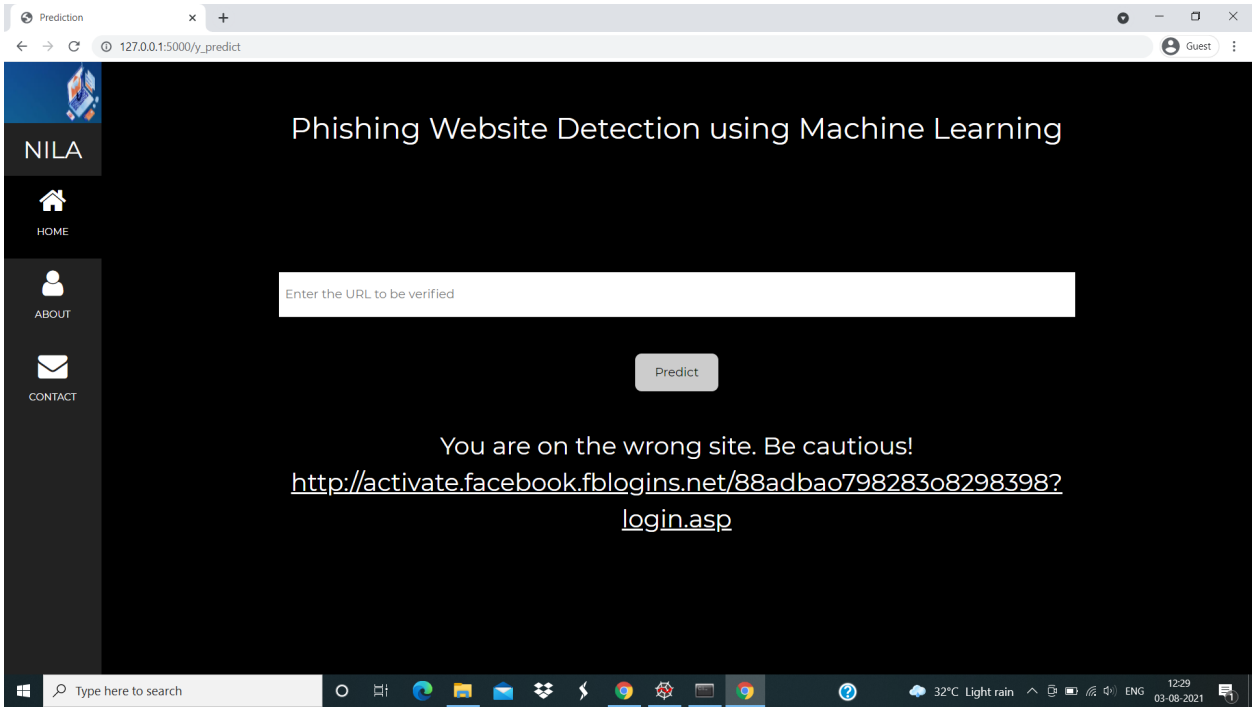
12. Appendix

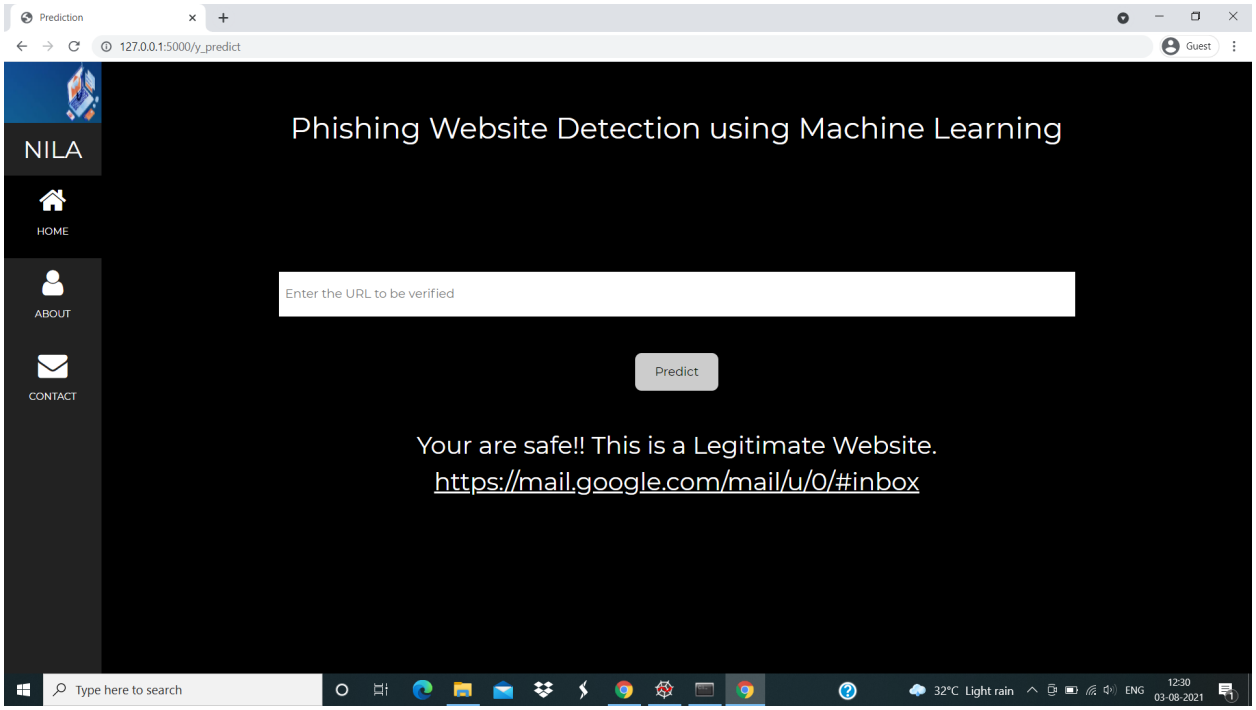
a. Source code -

github.com/smartinternz02/SI-GuidedProject-4913-1627463928

b. UI output Screenshot.


Prediction
127.0.0.1:5000/predict
Guest
NILA
HOME
ABOUT
CONTACT
Phishing Website Detection using Machine Learning
Enter the URL to be verified
Predict
Type here to search
32°C
ENG
12:27
03-08-2021


Prediction
127.0.0.1:5000/y_predict
Guest
NILA
HOME
ABOUT
CONTACT
Phishing Website Detection using Machine Learning
Enter the URL to be verified
Predict
You are on the wrong site. Be cautious!
http://activate.facebook.fblogins.net/88adbao798283o8298398?login.asp
Type here to search
32°C Light rain
ENG
12:29
03-08-2021


Prediction
127.0.0.1:5000/y_predict
Guest
NILA
HOME
ABOUT
CONTACT
Phishing Website Detection using Machine Learning
Enter the URL to be verified
Predict
Your are safe!! This is a Legitimate Website.
https://mail.google.com/mail/u/0/#inbox
Type here to search
32°C Light rain
ENG
12:30
03-08-2021