ISSN No:-2456-2165

Fraudulent Website Detection

[1] Guide – Vadipina Amarnadh (Assistant Professor)
[2] G Shreya ; [3] K Nikhil Chary; [4] N Naga Lakshmi
Department of CSE
Anurag Group of Institutions

Abstract:- Fraud refers to criminal deception that convinces victims to reveal personal information such as their password or credit card number. Fraudulent websites are usually designed to appear professional and convincing, as if they are genuine. To mitigate the negative effects of a fraudulent website. We proposed an effective phishing website detection system that analyses phishing website URL addresses in order to learn data patterns that can distinguish between authentic and phishing websites. To learn data patterns in website URLs, our system uses machine learning techniques such as decision trees. Using a random forest classifier, we evaluate our system on a recent phishing website dataset.

I. INTRODUCTION

When it comes to cyberattacks all over the world, phishing remains one of the most popular techniques that combines technical deception and social engineering to obtain information. The victim's personal information, such as login credentials, credit card details, or Organization sensitive information. Earlier in the Internet's history, blacklists were an effective way to keep track of fraudulent websites. However, phishing attackers are more than willing to flood the web with short-lived phishing attempts. The short lifespan of such malicious efforts can actually work in their favour because they rarely last long enough to make it onto the blacklist. This trend suggests that we should investigate more adaptable approaches. Detection and prevention of phishing websites continue to be a major area of study. There are various types of phishing techniques that provide torrential and essential ways for attackers to gain access to people's and organizations' data. With its ability to generalize and infer from past data, machine learning is one of the viable candidates for addressing this problem. The universal resource locator URLs, also known as "Web links," are critical components of a phishing attack. The vulnerability of the uniform resource locator is the ability to redirect the pages, i.e., through the hyperlink; which could redirect to the legitimate website or the phishing site. Every day, new methods for creating phishing sites emerge. This prompted several researchers to increase their focus on locating phishing sites.

II. EXISTING SYSTEM

A. Using site server log knowledge.

This technique for differentiating the examination of authentic site server log knowledge is required for phishing website. An off-the-shelf application or identification of a phishing website. Free, demonstrates a number of outstanding properties such as high precision, total autonomy, and nice language-freedom, speed of selection, flexibility to dynamic phish, and flexibility to advance in phishing methods.

B. Using PhishStorm.

PhishStorm is an automated phishing detection system that can analyse any URL in real time to detect potential phishing sites. To protect users from phishing content, Phish Storm is proposed as an automated real-time URL phishingness rating system. PhishStorm generates a phishingness score for each URL and can be used to rate the reputation of a website.

III. PROPOSED SYSTEM

The proposed model/system is a Fraudulent Website Detection with URLs using Machine Learning, which takes URL as input and classifies URLs using the Random Forest Algorithm (ML). We primarily used Five Features under the Feature Extraction module to determine whether or not the URL is legitimate. This system is simple to use, and the user can determine whether the website he or she is visiting is legitimate or not.

- *Random Forest Classifier*

A random forest is a meta-learner composed of multiple distinct trees. Each tree categorizes the data set overall, and the random forest approach selects the classification with the most votes. In other words, some entities will appear multiple times in the sample, while others will not. Each decision tree is built using a model based on a distinct random subset of the training dataset and a random subset of the available variables to identify how to partition the training dataset optimally at each node. The Random Forest decision tree models are combined to form the final ensemble model, with the majority determined by the votes of the individual decision trees.

IV. EXPERMENTAL TOOLS

- *SKLearn*

Scikit-learn provides a variety of supervised and unsupervised learning techniques via a standardised Python interface. It is available in several Linux distributions and is distributed under a liberal simplified BSD licence that encourages both academic and commercial use.

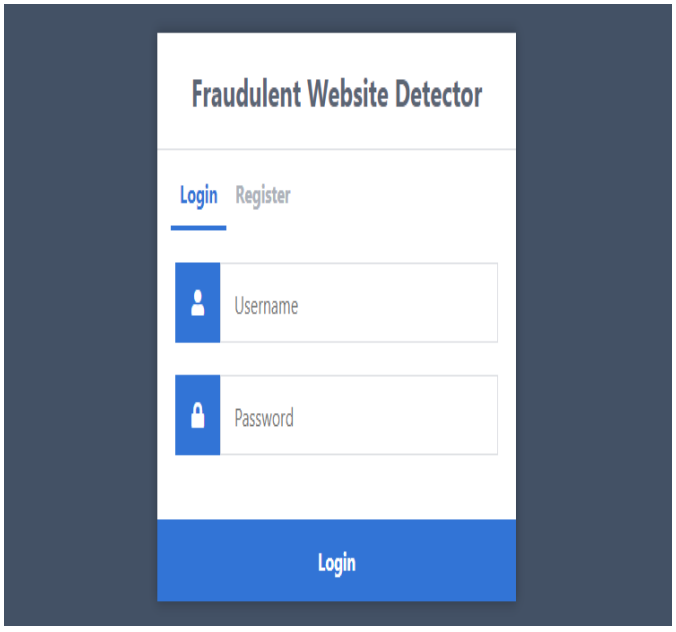
➢ *NumPy*

NumPy is a NumPy package in Python. The term "Numeric Python" or "Numerical Python" is an acronym. It is pronounced / (NUM-PY). It's a Python extension module written primarily in C. As a result, NumPy's precompiled mathematical and numerical functions and features guarantee fast execution. NumPy also extends the Python programming language with robust data structures such as multidimensional arrays and matrices. These data structures ensure that matrices and array calculations are accurate. The solution even aims for massive matrices and arrays, also known as "big data." Furthermore, the module includes a large library of complex mathematical operations that can be applied to these matrices and arrays.

➢ *Flask*

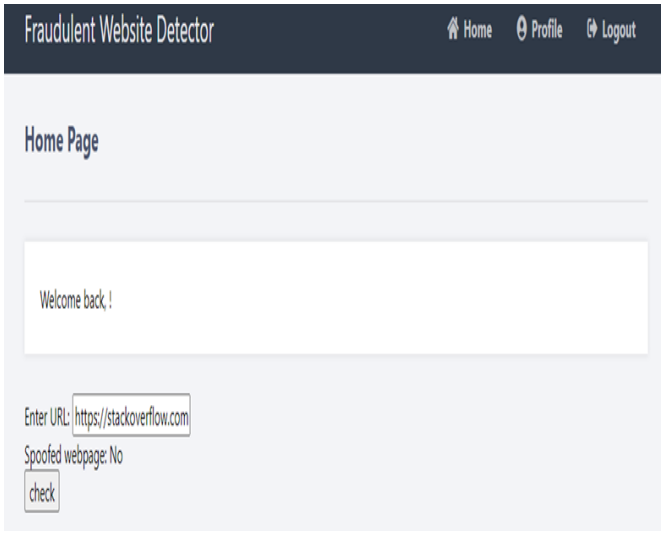
Flask is a web framework and a Python module that allows you to easily create web applications. It has a small and simple core: it's a microframework without an ORM (Object Relational Manager) or similar features. It has a lot of cool features, such as url routing and a template engine. It is a WSGI web application framework. Although it is a microframework, this does not imply that your entire app should be contained within a single Python file. To handle complexity, you can and should use multiple files for larger programs. The Flask framework is described as micro because it is simple but extensible. Flask does not make any decisions for you, such as which database to use or whether to use an ORM.

V. OUTPUT



(Fig.1 User Interface)

Users can register themselves by giving their name and email id and they can submit the website URL they are checking if it is not legitimate or not.



(Fig.2 example of a legitimate website)

(Fig.2 example of a spoofed website)

VI. CONCLUSION

The most important step in protecting the user is to educate them about phishing attacks.

Internet users must heed all security advice given by professionals. Furthermore, users should be taught not to click on links that take them to websites where they must submit sensitive information. The URL must be validated before accessing the website. It has been discovered that phishing attempts are extremely important, and it is critical that we develop a system to detect them. Addressing this issue is becoming increasingly important because phishing websites may reveal highly sensitive and private information about the user. Using a classifier and any machine learning technique will quickly and easily solve this problem. We already have classifiers that provide good phishing prediction rates, but our survey found that using a hybrid technique to make predictions and increase the accuracy of phishing website prediction rates would be preferable. Because the current method is less accurate, we created a new phishing technique that makes use of URL-based features.

REFERENCES

[1]. A Survey on Phishing And It's Detection Techniques Based on Support Vector Method (SVM) and Software Defined Networking(SDN). A. MahaLakshmi, N. Swapna Goud, Dr. G. Vishnu Murthy.
[2]. A Certain Investigations on Web Security Threats and Phishing Website Detection Techniques. N. Swapna Goud, Dr. Anjali Mathur.
[3]. Wong, R. K. K. (2019). An Empirical Study on Performance Server Analysis and URL Phishing Prevention to Improve System Management Through Machine Learning. In Economics of Grids, Clouds, Systems, and Services: 15th International Conference,GECON2018,Pisa,Italy,September 18-20, 2018, Proceedings (Vol. 11113, p. 199). Springer.
[4]. Marchal, S., Saari, K., Singh, N., & Asokan, N. (2016, June). Know your phish: Novel techniques for detecting phishing sites and their targets. In 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS) (pp. 323-333). IEEE.
[5]. https://www.proofpoint.com/us/threat-reference/phishing