# Web Phishing Detection

**Team Leader :** Rahul Haobijam (730919104085)

**Teem Members :** N. Chingkheinganba (730919104072)

Mutum Robert (730919104066)

Vishal Bhandari (730919104132)

# PRE-REQUISITE:

* Install Anaconda
* Install Python Packages
* Pandas
* Numpy
* Matplotlib
* Seaborn
* Flask
* Prior Knowledge

# GiTHUB ACCOUNT :

* Open [https://github.com](https://github.com/) in a web browser, and then select Sign up**.**
* Enter your email address.
* Create a password for your new GitHub account, and Enter a username**,** too. Next, choose whether you want to receive updates and announcements via email, and then select Continue**.**
* Verify your account by solving a puzzle. Select the Start Puzzle button to do so, and then follow the prompts.
* After you verify your account, select the Create account button.
* Next, GitHub sends a launch code to your email address. Type that launch code in the Enter code dialog, and then press Enter.
* I have created my github account with the email id : **hijoshua2001@gmail.com**

GitHub id: **IBM-Project-45782-1660732314**

# INSTALLATION OF IDE’S:

Python is available from its website python.org. Once there, hover your mouse over the Downloads menu, then over the Windows option, and then click the button to download the latest release.

# LITERATURE SURVEY:

**REVIEW- 1**

# Title Of The Paper:

# Detecting Phishing Website Using Machine Learning.

# Name Of The Author:

# Mohammed Hazim Alkawaz, Stephanie Joanne Steven, Asif Iqbal Hajamydeen.

# Problem Description:

Trying to gather personal information through deceptive ways is becoming more common nowadays. In order to assist the user to be aware of the access to such websites, the implemented system notifies the user through email and also pop-up, when trying to access a phishing site. This paper proposes an approach of phishing detection system to detect blacklisted URL also known as phishing websites, so that individual can be alerted while browsing or accessing a particular website. Therefore, it can be utilized for identification and authentication and become a legitimate tool to prevent an individual from getting tricked.

# REVIEW 2:

**Title Of The Paper** :

Phishing Website Detection using Machine Learning Algorithms.

# Name Of The Author:

# Rishikesh Mahajan, Irfan Siddavatam.

# Problem Description:

Phishing attack is a simplest way to obtain sensitive information from innocent users. Aim of the phishers is to acquire critical information like username, password and bank account details. Cyber security persons are now looking for trustworthy and steady detection techniques for phishing websites detection. This paper deals with machine learning technology for detection of phishing URLs by extracting and analyzing various features of legitimate and phishing URLs. Decision Tree, random forest and Support vector machine algorithms are used to detect phishing websites. Aim of the paper is to detect phishing URLs as well as narrow down to best machine learning algorithm by comparing accuracy rate, false positive and false negative rate of each algorithm.

# REVIEW 3:

**Title Of The Paper:**

Detecting phishing websites using machine learning technique.

# Name Of The Author:

Ashit Kumar Dutta, Dutta AK (2021) Detecting phishing websites using machine learning technique.

# Problem Description:

In recent years, advancements in Internet and cloud technologies have

led to a significant increase in electronic trading in which consumers make online purchases and transactions. This growth leads to unauthorized access to users’ sensitive information and damages the resources of an enterprise. Phishing is one of the familiar attacks that trick users to access malicious content and gain their information. In terms of website interface and uniform resource locator (URL), most phishing webpages look identical to the actual webpages. Various strategies for detecting phishing websites, such as blacklist, heuristic, Etc., have been suggested. However, due to inefficient security technologies, there is an exponential increase in the number of victims. The anonymous and uncontrollable framework of the Internet is more vulnerable to phishing attacks. Existing research works show that the performance of the phishing detection system is limited. There is a demand for an intelligent technique to protect users from the cyber-attacks. In this study, the author proposed a URL detection technique based on machine learning approaches. A recurrent neural network method is employed to detect phishing URL. Researcher evaluated the proposed method with 7900 malicious and 5800 legitimate sites, respectively. The experiments’ outcome shows that the proposed method’s performance is better than the recent approaches in malicious URL detection.

# REVIEW 4:

**Title Of The Paper:**

Phishing Website Detection Based on Machine Learning Algorithm.

# Name Of The Author:

# Weihengbai. Information Security Institute Johns Hopkins University Baltimore, USA.

# Problem Description:

Phishing websites are a means to deceive users' personal information by using various means to impersonate the URL address and page content of a real website. This paper analyzes the structural features of the URL of the phishing website, extracts 12 kinds of features, and uses four machine learning algorithms for training. Then, use the best performing algorithm as our model to identify unknown URLs. After the recognition is completed, a snapshot of the web page is extracted and compared with the regular web page snapshot to implement the recommendation of the original regular web page of the phishing web page.

# REVIEW 5:

**Title Of The Paper:**

Phishing Website Detection based on Supervised Machine Learning with Wrapper Features Selection.

# Name Of The Author:

# Waleed Ali, Department of Information Technology Faculty of Computing and Information Technology, King Abdulaziz University Rabigh, Kingdom of Saudi ArabiaProblem Description.

# Problem Description:

# The problem of Web phishing attacks has grown considerably in recent years and phishing is considered as one of the most dangerous Web crimes, which may cause tremendous and negative effects on online business. In a Web phishing attack, the phisher creates a forged or phishing website to deceive Web users in order to obtain their sensitive financial and personal information. Several conventional techniques for detecting phishing website have been suggested to cope with this problem. However, detecting phishing websites is a challenging task, as most of these techniques are not able to make an accurate decision dynamically as to whether the new website is phishing or legitimate. This paper presents a methodology for phishing website detection based on machine learning classifiers with a wrapper features selection method. In this paper, some common supervised machine learning techniques are applied with effective and significant features selected using the wrapper features selection approach to accurately detect phishing websites. The experimental results demonstrated that the performance of the machine learning classifiers was improved by using the wrapper-based features selection.The machine learning classifiers with the wrapper-based features selection outperformed the machine learning classifiers with other features selection methods.

# Paper Reference:

# [1] N. Abdelhamid, A. Ayesh, F. Thabtah, “Phishing detection based associative classification data mining,” Expert Systems with Applications, vol. 41(13), pp. 5948-5959, 2014.

# [2] R. M. Mohammad, F. Thabtah, L. McCluskey, “Tutorial and critical analysis of phishing websites methods,” Computer Science Review, vol. 17, pp. 1-24, 2015.

# [3] H. Huang, S. Zhong, J. Tan, “Browser-side countermeasures for deceptive phishing attack,” Fifth International Conference on Information Assurance and Security IAS'09, vol. 1, pp. 352-355, IEEE, 2009.

# [4] R. M. Mohammad, F. Thabtah, L. McCluskey, “Predicting phishing websites based on self-structuring neural network,” Neural Computing and Applications, vol. 25(2), pp. 443-458, 2014.

# [5] M. A. U. H. Tahir, S. Asghar, A. Zafar, S. Gillani, “A Hybrid Model to Detect Phishing-Sites Using Supervised Learning Algorithms,” International Conference on Computational Science and Computational Intelligence (CSCI), pp. 1126-1133, IEEE, 2016.

# [6] M. He, S.J. Horng, P. Fan, M.K. Khan, R.S. Run, J.L. Lai, R.J. Chen, Sutanto, “An Efficient Phishing Webpage Detector,” Expert Systems with Applications, vol. 38(10), pp. 12018-12027, 2011.

# [7] P. A. Barraclough, M. A. Hossain, M. A. Tahir, G. Sexton, N. Aslam, “Intelligent Phishing Detection and Protection Scheme for Online Transactions,” Expert Systems with Applications, vol. 40(11), pp. 4697- 4706, 2013.

[8] H. H. Nguyen, D. T. “Nguyen, Machine learning based phishing web sites detection,” In AETA 2015: Recent Advances in Electrical Engineering and Related Sciences , pp. 123-131, Springer International Publishing, 2016.

[9] R. M. Mohammad, F. Thabtah, L. McCluskey, “Intelligent Rule-based Phishing Websites Classification, “ IET Information Security, vol. 8(3), pp. 153-160, 2014.

[10] V. S. Lakshmi, M. S. Vijaya, “Efficient prediction of Phishing Websites Using Supervised Learning Algorithms,” Procedia Engineering, vol. 30, pp. 798-805, 2014