LITERATURE SURVEY-1

TOPIC: Phishing Website Detection Using Machine Learning.

AUTHORS: AdarshMandadi;SaikiranBoppana;Vishnu Ravella;R Kavitha.

YEAR: 2022

Phishing is an **Internet scam** in which an **attacker sends fake messages that appear to come from a trusted source**. The **goal** was to **get as many people as possible to click on a link or open an infected file**. There are different approaches to detect this type of attack. **One approach is machine learning**. The **URLs received from the user are fed into the machine learning model**, then the **algorithm processes the input and displays the output**, whether it is **phishing or legitimate**. The **proposed approach deals with Random Forest, Decision Tree classifiers**. The proposed approach effectively classified phishing and legitimate URLs with an **accuracy of 87.0% and 82.4%**, respectively, for the Random Forest and Decision Tree classifiers.

LITERATURE SURVEY-2

TOPIC: An Ensemble Method for PhishingWebsites Detection Based on XGBoost.

AUTHORS: Jiaqi Gu;HuiXu.

YEAR:2022

Today, the **Internet is spreading all over the world**. Apart from the benefits that the Internet brings to its users, there are many potential harms such as **phishing scams**.In this paper, I present an ensemble model for detecting phishing websites using URL functions. I used the **dataset called "Phishing Website Detector - Phishing Website Dataset" from Kaggle**. Then, several models were built using all sorts of common ensemble algorithms (like stacking, boosting, and bagging). Finally, various methods of measuring the performance of the models have been used. The **reason why only ensemble models** are chosen as methods to implement is mainly because of their **good overall performance**. In terms of results, the proposed **XGBoost model combining Random Forest and K-Nearest Neighbors**outperforms all other models (Random Forest, AdaBoost, Parameterized XGBoost, Stacking and Voting), with **an accuracy of 99.74% on data from training and 96.44% on test data.**

LITERATURE SURVEY-3

TOPIC: Phishing Detection and Prevention using Chrome Extension.

AUTHORS:M. Amir SyafiqRohmat Rose ; NurlidaBasir ; NurFatin Nabila RafieHeng ; Nurzi Juana MohdZaizi ; MadihahMohd Saudi.

YEAR:2022

During the COVID-19 pandemic outbreaks, the number of cyber attacks, including phishing activities, skyrocketed. Many **technical solutions for phishing detection have been developed today**, but these approaches have failed or failed to identify phishing pages and efficiently detect malicious code. One of the **drawbacks is poor detection accuracy and poor adaptability to new phishing connections**.In this paper, an intelligent phishing detection and prevention model is designed. The **proposed model uses a self-destruct detection algorithm** in which machine learning, especially **supervised learning algorithm, has been used**. **All rules** used in the algorithm **focus on URL**-based web properties that attackers rely on to redirect victims to the simulated websites. Accordingly, based on the proposed model, **Phishing Detection for Chrome Extensions** has been developed to **prevent phishing attacks** with appropriate **countermeasure and alert users** of phishing when visiting illegal websites. This smart phishing detection and prevention model is believed to be able to prevent scam and spam websites and reduce cybercrime and cybercrisis happening year by year.

LITERATURE SURVEY-4

TOPIC: Phishing Web Page Detection Methods: URL and HTML Features

Detection.

AUTHOR: Human Faris,SetiadiYazid.

YEAR:2021

Phishing is a type of fraud on the Internet in the **form of fake web pages** that mimic the original web pages to trick users into sending sensitive information to phisher. The statistics presented by APWG and Phistank show that the number of phishing websites from 2015 to 2020 tends to increase continuously. To

overcome this problem, several studies have been carried out including detecting phishing web pages using various features of web pages with various methods. Unfortunately, the use of several methods is not really effective because the design and evaluation are only too focused on the achievement of detection accuracy in research, but evaluation does not represent application in the real world. Whereas a security detection device should require effectiveness, good performance, and deployable. In this study the authors evaluated several methods and proposed rules-based applications that can detect phishing more efficiently.

LITERATURE SURVEY-5

TOPIC:A Lightweight Phishing Website Detection Algorithm by Machine Learning.

AUTHOR: ChenyuGu.

YEAR:2021

With the rapid development of the Internet, phishing websites now show the characteristics of short life cycle and low construction cost, which leads to a large amount of data brought by the **detection of phishing websites for URL** (uniform resource locator). It will also lead to increased retrieval time and decreased detection speed. In order to deal with diverse, complex and hidden phishing websites, this paper proposes a lightweight framework for detecting phishing websites. We first choose **the faster Minhash signature to match URLs**. On one hand, similarity detection is employed if the websites is suspicious. On the other hand, based on machine learning, the phishing websites can be finally determined by **intention detection without similar sites**.

LITERATURE SURVEY-6

TOPIC:A Review on Phishing Websites Revealing through Machine Learning.

AUTHOR: Alok Singh Sengar; AbhishekBhola; Ratnesh Kumar Shukla;Anurag Gupta.

YEAR:2021

Phishing is a frequent assault in which unsuspecting people's unique, private, and **sensitive information is stolen through fake websites**. The primary objective of phishing websites consistent resource allocators is to steal unique, private, and sensitive information such as user login passwords and online financial transactions. Phishers construct phony websites that look and sound just like genuine things. With the advent of technology, **there are protecting users significantly increased in phishing methods.** It necessitates the development of an anti-phishing technology to identify phishing and protect users. **Machine learning is a useful technique for combating phishing attempts**. These articles were utilized to examine Machine learning for detection strategies and characteristics.

LITERATURE SURVEY-7

TOPIC: Fishing out the Phishing Websites

AUTHORS: Peya mowar, mini jain

YEAR: 2021

Phishing is a type of cybercrime when fraudulent websites entice unsuspecting visitors and coerce them into divulging private information like financial or social network passwords. Phishing websites are designed **to superficially resemble well-known trustworthy websites.** This study proposes a unique classifier that takes into account lexical-based, script-based, rule-based, and address-based data collected from a website in order to identify such phishing websites. The XGBoost (eXtreme Gradient Boosting) model outperforms all other tree-based ensemble classifiers with a testing accuracy of 99.6% using a large-scale balanced dataset of 38,800 active phishing and legal websites. Without relying on external **factors like page rank, the classifier is capable of identifying zero-day phishing assaults**.

LITERATURE SURVEY-8

TOPIC: Phishing Detection from URLs Using Deep Learning Approach

AUTHORS:Shweta Singh; M.P. Singh; Ramprakash Pandey

Year:2020

The Internet is now accessible everywhere. People enjoy using an online marketplace to buy or sell their goods all over the world. As a result, cyberattackers now gravitate toward crimes in cyberspace. Phishing is one such strategy where attackers/criminals have used **the unidentified structure of the Internet** with the intention of misleading people with the use of the **fictitious website and emails in order to gain their credentials** (like account numbers, passwords, and PINs). Due to this semantic framework, it might be difficult to tell whether a web page is real or phishing. In order to stop such attempts, a **phishing detection system is created in this work using deep learning methods**. The method utilises a convolutional neural network (CNN) to analyse URLs in order to identify the phishing webpage.

LITERATURE SURVEY-9

TOPIC: HTMLPhish: Enabling Phishing Web Page Detection by Applying Deep Learning Techniques on HTML Analysis

Author: Chidimma Opara; Bo Wei; Yingke Chen

Year: 2020

The creation and execution of phishing attacks now only demand minimal technical expertise and expense. This rebellion has resulted in an increase in phishing attacks on the Internet. As a result, proactive methods of thwarting **phishing assaults are now absolutely essential**. In this research, we offer HTMLPhish, **a data-driven end-to-end automatic phishing web page classification solution that is based on deep learning**. In particular, HTMLPhish uses convolutional neural networks (CNNs) to learn the semantic dependencies in the textual contents of the HTML document after receiving the content of the HTML document from a web page. Without considerable human feature engineering, **the CNNs learn appropriate feature representations from the HTML page embeddings**. Additionally, the concatenation of the word and character embeddings in our suggested method enables our model to handle new features and provide simple extrapolation to test data. On a dataset

of more than 50,000 HTML documents, we undertake extensive trials that produce results with a true positive rate and accuracy of over 93%. This dataset includes a distribution of phishing to safe web pages that are accessible in the actual world. Additionally**, HTMLPhish is a client-side method** that is totally language-independent, enabling it to undertake web page phishing detection regardless of the **textual language**.

LITERATURE SURVEY-10

TOPIC: On Effectiveness of Source Code and SSL Based Feature for Phishing Website Detection.

Author: S. Roopak; Athira P. Vijayaraghavan; Tony Thomas

Year: 2019

A social engineering technique called phishing is used to acquire user credentials from rogue websites' data entry forms. Only phishing websites that are blacklisted by **anti-malware programmes** can be found nowadays. By making modest adjustments to the textual and visual contents of a phishing site, it is simple to avoid **similarity-based detection techniques** like visual similarity. From a web page's URL, domain, and source code-based characteristics, phishing behaviour can be recognised. By utilising **black hat** SEO strategies, URL- and domain-based functionality can be quickly bypassed. In this research, we use the **Repeated Incremental Pruning** to Produce Error Reduction (RIPPER) technique to extract the pertinent rules from a training dataset based on Secure Socket Layering (SSL) based characteristics and webpage source code.