LITERATURE SURVEYS

Domain	Machine Learning
Project name	Web Phishing Detection
Team id	PNT2022TMID51354

LITERATURE SURVEY - I

PHISHING DETECTION USING MACHINE LEARNING ALGORITHMS

Phishing has been characterized as a social assault that uses several platforms to obtain information from web users. Phishing attacks are one of the most serious hazards to individuals and businesses in today's digital world. Many attacks are performed each month with the goal of convincing consumers that they are visiting a reputable website or online application in order to obtain account information. A strategy is provided for detecting these types of assaults by modifying existing Document Object Model (DOM) comparison tools such as Proportional Distance, false positive and false negative, and the favicon image recognition algorithm. There may be a different impact, misclassifying a phishing website than misclassifying a legitimate website if the proportional distance approach is used. To solve this, the proportional distance method is used. The goal is to develop a scheme which protects user from phishing attacks. The interface includes three pre-existing methods which used to detect the phishing attacks

Proportional Distance

false positive and false negative

Favicon images recognition algorithm

The characteristics of phishing assaults is described and a categorization model is provided for phishing attacks. This approach entails extracting features from webpages as well as a categorization part.

LITERATURE SURVEY – II

DETECTING PHISHING WEBSITES USING MACHINE LEARNING

Phishing website is one of the internet security problems that target the human vulnerabilities rather than software vulnerabilities. It can be described as the process of attracting online users to obtain their sensitive information such as usernames and passwords. This paper offered an intelligent system for detecting phishing websites. The system acts as an additional functionality to an internet browser as an extension that automatically notifies the user when it detects a phishing website. The system is based on a machine learning method, particularly supervised learning. Random Forest technique was selected due to its good performance in classification. The focus was to pursue a higher performance classifier by studying the features of phishing website and choose the better combination of them to train the classifier. As a result, the paper is concluded with accuracy of 98.8% and combination of 26 features.

All 36 features were studied in order to reduce time computation and providing high performance with the least combination of the powerful features. Because of time shortage and hardware limitation, random features were chosen to process its combination. After some observation it concluded that the combination of features computed take the shape of normal distribution curve, it starts with least combination of features with low probability of combination and time consuming, then picks up accordingly, then goes down as it reaches final number of 36 features

LITERATURE SURVEY – III

PHISHING DETECTION FROM URLs USING DEEP LEARNING APPROACH

Phishing is one such technique where the unidentified structure of the Internet has been used by attackers that intend to deceive users with the use of the illusory website and emails for obtaining their credentials (like account numbers, passwords, and PINs). Consequently, the identification of a phishing or legitimate web page is a challenging issue due to its semantic structure. In this paper, a phishing detection system is implemented using deep learning techniques to prevent such attacks. The system works on URLs by applying a convolutional neural network (CNN) to detect the phishing webpage. This system doesn't require any feature engineering as the CNN extract features from the URLs automatically through its hidden layers. This is other advantage of the proposed system over earlier reported in as the feature engineering is a very time-consuming task.

To detect the phishing URLs in real-time, the proposed phishing detection system firstly, tokenize the URLs that pass it through the glove embedding layer and then it is passed through CNN Deep neural network, followed by the dense neural network to gain the F-1 score precision, recall, accuracy, support of the legitimate and phishing URLs. The proposed model-based network can classify URLs. The result shows that the system can identify phishing URLs with an accuracy of 98.00 percent.

LITERATURE SURVEY – IV

DETECTING PHISHING WEBSITE USING MACHINE LEARNING

Phishing is the act of pretending to be a legitimate website in order to deceive users into giving up their personal information, such as usernames, passwords, account numbers, social security numbers, etc. Possibly the most pervasive cybercrime committed today is phishing fraud. Numerous websites, including those that handle online payments, web-mail, financial institutions, file hosting or cloud storage, and many others, are susceptible to phishing attacks. More than any other industry area, the web-mail and online payment sector was threatened by phishing. Since spear phishing and email phishing scams can both be used to commit phishing, users should be aware of the risks and not place their complete reliance in conventional security software.

The installed system alerts the user by email and pop-up while attempting to enter a phishing site in order to let the user be aware of the access to such websites. In order for a person to be warned while browsing or accessing a particular website, this paper provides a method of phishing detection system to identify blacklisted URLs, also known as phishing websites. It may be used for identification and authentication and develop into a reliable tool to stop someone from being fooled. This system can be improved by sending real-time notification when a phishing website is accessed.

LITERATURE SURVEY – V

A New Method for Detection of Phishing Websites: URL Detection

In this paper, a novel method to detect website phishing is developed using the Random Forest as classification algorithm with the help of RStudio. Phishing is an illegal activity where people are misled into the wrong sites by using various fraudulent methods. These phishing websites confiscate private information or financial details for ill intentions. The model proposed in this paper uses the URL detection method using Random Forest algorithm to detect phishing websites. There are three major phases in implementing the proposed model namely,

1. Parsing
2. Heuristic Classification of Data
3. Performance Analysis

This model has used a wide range of metrics, including true positives, true negatives, false negatives, the F-measure, ROC, precision, and sensitivity for analysis purposes thus giving a clear view on the performance and accuracy each time the detection takes place.