Date	19 September, 2022
Team ID	PNT2022TMID37820
Project Name	Web Phishing Detection
Maximum Marks	4 Marks

Web Phishing Detection

Abstract

Phishing is a attack against internet users that causes them to reveal their info using fake websites. The goal of the fake website is to steal information as usernames, passwords and online banking transactions.

Machine learning is a powerful tool used to combat spoofing attacks. This reports covers machine learning applied science to detect fake URLs by extracting and analyze different feature of legitimate and fake URLs. Logistic Regression and algorithms are use to detect fake websites.

Introduction

Nowadays the Internet plays an important role in communication, where an online environment to manage business, social networks. However the Internet also contains hidden things a lot of risk cause when users operate in an online environment they can vulnerable to attackers. Often place on popular websites or sent to user emails.

Literature Survey

1. Web phishing detection using deep learning framework

Despite numerous research efforts, phishing attacks remain prevalent and effective in unsuspecting users to reveal sensitive information including social security numbers. In this project we propose Phishing a new feature rich machine learning framework to detect phishing webpages. These features capture various characteristics of legitimate web applications as well as their underlying web infrastructures. In our experiments, Phishing achieved 91.6% accuracy with 8.4% false positive rate on a dataset containing unique phishing instances using machine learning framework.

2. Spam detection using machine learning techniques

Nowadays email are use in almost every field , from business to education. Emails has two subcategories, ham and spam. Also call junk email is a type of email that can be use to harm any user by stealing valuable info. Spam detection is significant and enormous problem for IOT service providers nowadays. Preventing and filtering is essential approaches. Several machine learning and deep learning technique have been use to get comprehensive comparison of these technique to get base on accuracy and precision are discuss.

3. A machine learning based approach for phishing detection using hyperlinks information

Phishing is a kind of attack in which perpetrators use spoofed emails and fraudulent web sites to lure unsuspecting online users into giving up personal info. This project looks at the phishing problem by analyze various research works and counter measures, and how to increase detection.It consists of three studies. In the first study, focus was on the dataset gathering, pre processing features extraction and data set for the classification process. In the second study, focus on the metrics evaluation of a set of classifier using accuracy and precision. The second part of result outcome of the study shows the individual classifier method performed better with an accuracy of 91.6% while the result of the small size of dataset used as it is show in past researches, that K-NN performe better with a decreasing the size of dataset while classifier like SVM and C4.5 performed better with increasing size of dataset.

4. Deep learning for phishing detection : Taxonomy current challenges and future directions

Internet applied science is so pervasive today, for example from social networking to online banking it has made people's lives more easy. Due to applied science advancement , security threats to systems and networks are relentlessly inventive. one such a serious threat is "phishing", in which attackers attempt to steal the user credentials using fake emails or websites. It is true that both industry and academia are working hard to develop solutions to combat against phishing threats. It is therefore very important that organizations to pay attention to user end awareness In phishing threat prevention. Therefore, aim is to protects the users from phishing. Based on the attacks, we conclude the issues and challenges that still exists in the literature, which are important to fight against phishing threats.

5. Machine learning based phishing detection from URL's

In the recent years, advancements in internet and cloud applied science have led to a significant increase in electronic trading in which consumers make online purchases and transactions. Phising is one of the similar attacks that trick users to access malicious content and gain their info. In terms of website interface and uniform resource locator most phishing websites, such as blacklist , heuristic have been suggest. However, due to the inefficient security applied science, there is an exponential increase in the number of victims. The anonymous and uncontrollable framework of the internet is more vulnerable to phishing attacks. A recurrent neural network method is employee to detect phishing URL. Researcher evaluated the proposed with 7900 malicious and 5800 legitimate sites, respectively. The experiments outcome shows the better performance than the recent approaches in malicious URL detection.

Conclusion

This paper aims to enhance detection method to detect phishing websites using machine learning technology. We achieved 97.14% detection accuracy using random forest algorithm with lowest false positive rate. Also result shows that classifiers give better performance when we used more data as training data.