PLAGIARISM SCAN REPORT

Date May 08, 2021

Exclude URL: NO

Unique Content	97%
Plagiarized Content	3%
Paraphrased Plagiarism	0

Word Count	1,233
Records Found	3

CONTENT CHECKED FOR PLAGIARISM:

A Project reportonIntelligent Rule Based Phishing Website DetectionSubmittedByName of students (RollNo.)Anjuru Lokesh - 18BCS006A.Sai Venkata Aditya - 18BCS011CH.N.V.Avinash - 18BCS021D.Geetha Krishna - 18BCS025Under the guidance ofUMA SINDIAN INSTITUTE OF INFORMATION TECHNOLOGYDHARWAD Table of Contents1.INTRODUCTION2 .DATASET3.FEATURE EXTRACTION4.RANDOM FOREST ALGORITHM5.IMPLEMENTATION AND RESULT6. INTEGRATED PROJECT WITH DATABASE7.WORKING OF THE PROJECTIndian Institute of Information Technology, Dharwad ABSTRACT :This project aims to use an intelligent, versatile,and reliable method to recognise andclassify e-banking and other phishing websites. Inthis project, we check if a given URL is Phishy ornot by looking it up in a dataset, and if it's notthere, we'll break it down based on factors liketheURL, Domain Identity and Security, and encryptionrequirements, and determine whether it's Phishyor not. Phishing websites are those that ask usersfor personal information for a malicious reason.OBJECTIVE:Web phishing is one of many security threats to webservices on the Internet. Webphishing aims to steal private information, such asusernames, passwords, and credit carddetails, by way of impersonating a legitimate entity.So this project mainly focuses onapplying a machine learning framework to detect phishingwebsites.1. INTRODUCTIONPhishing has become a major source of concern forsecurity researchers in recentyears because it is relatively simple to create afake website that appears to be identical to alegitimate website. While experts can recognise fakewebsites, not all users can, and as aresult, some users become victims of phishing attacks.The attacker's main goal is to stealbank account credentials. Because of a lack of userknowledge, phishing attacks arebecoming more successful. Since phishing attacks

takeadvantage of user flaws, it's difficultto minimise them, but it's critical to improve phishingdetection techniques.The "blacklist" approach is a general method fordetecting phishing websites byadding blacklisted URLs and IP addresses to the antivirusdatabase. To get around blacklists,attackers use clever techniques like obfuscation andmany other simple techniques likefast-flux, where proxies are automatically createdto host the web page; algorithmicgeneration of new URLs; and so on. The method's biggestflaw is that it can't detect zero-hourphishing attacks.Heuristic-based detection, which uses characteristicsthat have been observed inreal-world phishing attacks and can detect zero-hourphishing attacks, but the characteristicsare not guaranteed to always be present in such attacks,and the false positive rate in detectionis high. Indian Institute of Information Technology, Dharwad Many security researchers are now focusing on machine learning techniques to solvethe limitations of blacklist and heuristic-based methods. Machine learning technology ismade up of a variety of algorithms that use historicaldata to make predictions about futuredata. The algorithm can evaluate various blacklistedand legitimate URLs and their featuresin order to accurately detect phishing websites, includingzero-hour phishing websites, usingthis technique.2. DATASET :URLs of benign websites were collected from www.kaggle.comand The URLs ofphishing websites were collected y, Shortening Services like “TinyURL”:TinyURLservice helps phishers tomask long phishing URLs by shortening them. The aimis to guide users to phishing websites. If theURL is created using a URL shortening service (suchas bit.ly), function is set to 1, otherwise it isset to 0.9) Length of Host name:The average length of safeURLs is discovered to be a 25. If theURL's length is greater than 25, the function is setto 1 otherwise to 0.10) Presence of sensitive words in URL:Phishingwebsites use sensitive words intheir URLs to make users believe they are visitinga legitimate website. The following words can beused in several phishing URLs: - 'confirm', 'account','banking','secure', 'ebayisapi', 'webscr', 'signin', 'mail', 'instal', 'toolbar', 'backup', 'paypal','password', 'username', and so on;Indian Institute of Information Technology, Dharwad 11) Number of slash in URL:The number of slashes in benevolent URLs is found to be a5; if the number of slashes in the URL is greaterthan 5, the feature is set to 1; if the number ofslashes in the URL is less than 5, the feature isset to 12) Age of SSL Certificate:The presence of HTTPSis critical in giving the impressionthat a website is legitimate. However, the SSL certificateof a benign website should be between 1and 2 years old.13) URL of Anchor:By crawling the source code ofthe URL, we were able to extractthis function. The a> tag specifies the anchor's URL.The function is set to 1 if the a> tag has amaximum number of hyperlinks from another domain,otherwise to 0.14) Website Rank:We took the rank of the websitesand compared it to the first 100,000websites in the Alexa database. If the website's rankis

greater than 10,0000, the function is set to 1;otherwise, it is set to 0.4.Random Forest Algorithm :The random forest algorithm, which is built on theprinciple of the decision tree algorithm, isone of the most efficient algorithms in machine learningtechnology. The random forest algorithmgenerates a forest containing a large number of decisiontrees. A large number of trees results in ahigh level of detection precision.The bootstrap method is used to build trees. To builda single tree, the bootstrap methodselects features and samples from the dataset at randomwith replacement. Random forest algorithm,like decision tree algorithm, chooses the best splitterfor classification from randomly chosenfeatures. Random forest algorithm also uses gini indexand knowledge gain methods to find the bestsplitter. This process will be repeated until therandom forest has produced n trees.Indian Institute of Information Technology, Dharwad Each tree in the forest predicts the target value, and the algorithm then calculates the votesfor each target value predicted. Finally, the randomforest algorithm uses the projected target withthe most votes as a final prediction.5. IMPLEMENTATION AND RESULT :Machine learning algorithms were imported using theScikit-learn platform. In 50:50, 70:30,and 90:10 ratios, the dataset is divided into trainingand testing sets. Each classifier is trained usingatraining set, and the output of the classifiers isevaluated using a testing set. The accuracy score,false negative rate, and false positive rate of classifierswere calculated to assess their efficiency.6. INTEGRATED PROJECT WITH DATABASE :We incorporated a MongoDB database into our projectso that the URL is first checkedagainst the database. If the URL cannot be locatedin the database, it is separated into processes andthe ml algorithm is used to predict the outcome. Theuser is routed to the website if it is not phishy.If the URL is phyishy, it is immediately added tothe database and the user is informed. It will locatethe URL more reliably and effectively this way becauseit will not divide it further if it is found inthe database.7. WORKING OF PROJECT :● We will have an interface build which will takean URL as an input.● We will have a database in which all the Phishywebsites found until now will bepresent.● We will first check whether the input URL is presentin the database or not.● If the URL is present in the database then we willnot redirect the user into thatURL.Indian Institute of Information Technology, Dharwad ● And also we will display a message to the user thatthe particular URL is phishy.● If the URL is not present in the database then wewill divide the URL into 6 parts.● Each part of the URL is checked through some stepsand given some values as anoutput in each step.● We had a ML model which was trained using some dataset.● This ML model will predict whether the website isphishy or not based on thevalues we got above.● And if this ML model classifies the input URL asa phishy website then we will addthe URL into our database automatically.● We will not allow the user to redirect into thewebsite if the URL is phishy

andnotifies the user with a notification.● We will redirect the user into the website if it'snot phishy.8. TO TEST OUR PROJECT :To try out our project, go to the link below, whereyou'll find all of thespecifications and a readme that will guide you throughthe process.Github LinkIndian Institute of Information Technology, Dharwad

MATCHED SOURCES:

www.solutionessays.com - *1% Similar*Compare

https://www.solutionessays.com/

---

core.trac.wordpress.org - *1% Similar*Compare

https://core.trac.wordpress.org/report/31?sort=owner&asc....

---

docs.microsoft.com - *<1>Compare*

https://docs.microsoft.com/en-us/sql/integration-services/in....
