Literature Survey

Web Phishing Detection

TEAM ID: PNT2022TMID37408

TEAM MEMBERS:

Team leader : Mohamed Abul faiz

Team Member 1: Nowfal

Team Member 2: Syed Suheab

Team Member 3: Salamath

S. NO	TITLE	AUTHOR & YEAR	PROPOSED WORK	TOOLS USED / ALGORITHM	TECHNOLOGY	RESULTS	FUTURE SCOPE
1	Survey of review spam detection using machine learning tchniques	**AUTHOR:** Michael Crawford, Taghi M. Khoshogoftaar, joseph D , Prusa , Aaron N. Richter & Hamzah Al Najada . **YEAR:** 05 October 2015	Reliance on online reviews gives to the potential concern that wrongdoers may create false reviews to artificially promote or devalue products and services. This practice is known as opinion (Review) spam, where spammers manipulate and poison reviews (i.e., making fake, untruthful, or deceptive reviews) for profit or gain.	-Random Forest Algorithm. -Support Vector Machine Algorithm	Machine Learning	**ADVANTAGES:** ❖ In recent years, review spam detection has received significant attention in both business and academia due to the potential impact fake reviews can have on customer behaviour and purchasing detection. **DISADVANTGES**: ❖ Although there are a large number of machine learning algorithms(learners) available, current research using supervised learning methods has been, for the most part, limited to three learners: Logistic Regression (LR), Navie Bayes (NB) and support Vector Machine (SVM).	According to our survey they done a project only for the Review spam . But we are going to detect the whole phishing website.

S.NO	TITLE	AUTHOR & YEAR	PROPOSED WORK	TOOLS USED / ALGORITHM	TECHNOLOGY	RESULTS	FUTURE SCOPE
2	A Survey and classification of web Phishing detection schemes	**AUTHOR:** Manoj Misra, Pradeep K. Atrey. **YEAR:** 26, October 2016	Phishing is a fraudulent techniques that is used over the internet to device users with the goal of extracting their personal information such as username, passwords, credit card, and bank account information. The key to phishing is deception. Phishing uses email spoofing as its initial medium for deceptive communication followed by spoofed websites to obtain the needed information from the victims. This paper studies ,analyses, and classifies the most ignificant and novel strategies proposed in the area.		➢ Search engine-based technique. ➢ Heuristics and machine learning based technique . ➢ Phishing blacklist and whitelist-based technique . ➢ Visual similarity- based techniques . ➢ DNS-based techniques	**ADVANTAGES:** ❖ The paper focuses on the fact that phishing detection schemes perform better than phishing prevention and user training solutions because they do not require changes in authentication platforms and do not rely on the user's ability to detect phishing **DISADVANTGES**: ❖ If a webpage is carefully designed by Phisher the extracted features might not give enough information to detect Phishing ❖ It is difficult to detect phishing websites from their visual appearance or via security indicators on mobile phones due to their small screen size	According to our survey, we are going to work for giving it 100% Detection of sites.

S.NO	TITLE	AUTHOR & YEAR	PROPOSED WORK	TOOLS USED / ALGORITHM	TECHNOLOGY	RESULTS	FUTURE SCOPE
3	Phishing Website Detection Using Machine Learning Algorithms.	**AUTHOR:** Rishikhesh Mahajan, Irfan siddavatam. **YEAR:** October 2018	URLs extracting and analyze Various link by check with Backlisting with help of Machine Learning to increase accuracy.	• Decision Tree Algorithm • Random Forest Algorithm • Support Vector Machine Algorithm	• Machine Learning	**ADVANTAGES:** ❖It is 97.14% detection accuracy using random forest algorithm with lowest false positive rate. **DISADVANTGES**: ❖It is that the Characteristics are not guaranteed to always exist in such attacks and false positive rate in detection is very high . ❖It is also has a Negative effects on a business, including loss of money, loss of intellectual property	We are trying to reduce the false positive rate in detection.

S.NO	TITLE	AUTHOR & YEAR	PROPOSED WORK	TOOLS USED / ALGORITHM	TECHNOLOGY	RESULTS	FUTURE SCOPE
4	Intelligent web-phishing detection and protection scheme using integrated features of Images, frames and text.	**AUTHOR:** M.A.Adebowale , K.T.Lwin , E.Sanchez , M.A.Hossain . **YEAR:** January 2019	A phishing attack is one of the most significant problems faced by online users because of its enormous effect on the online activities performed. In recent years, phishing attacks continue to escalate infrequency, severity and impact. Several solutions, using various methodologies, have been proposed in the literature to counter the web-phishing threats.Not withstanding, the existing technology cannot detect the new phishing attacks accurately due to the insufficient integration of features of the text, image and frame the evaluation process.	• Decision Tree Algorithm • Random Forest Algorithm.	Machine Learning	**ADVANTAGES:** ❖ Adaptive Neuro-Fuzzy Inference System based robust scheme provide more accuracy. ❖ Combine features text, images & frames for phishing detection proof more detection. ❖ This is the first work that reflects the best unified text, image and frame feature. **DISADVANTGES**: ❖ Rules are generated by Neuro-Fuzzy logic are completely in agreement with the findings based on statistical analysis. ❖ The structure is not total interpretable.	We are trying to make a structure totally interpretable.

S.NO	TITLE	AUTHOR & YEAR	PROPOSED WORK	TOOLS USED / ALGORITHM	TECHNOLOGY	RESULTS	FUTURE SCOPE
5	Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning.	**AUTHOR:** Nguyet Quang Do, Ali Selamat, Ondrej Krejcar, Enrique Herrera-Viedma. **YEAR:** January 2020	It proposes an integrated phishing website detection method based on convolutional neural networks (CNN) and random forest (RF).	• Linear Regression • K nearest neighbour • Support Vector Machine • Random Forest • XG Boost • Naïve Bayes • RNN Model • CNN Model	-Machine Learning - Deep Learning.	**ADVANTAGES:** ❖Advantage is that the Third-party service is independent. ❖ There is no standard guideline for an optimal set of parameters that can produce the best performance accuracy. **DISADVANTGES**: ❖Disadvantage is that the model cannot determine whether the URL is active or not, so it is necessary to test whether the URL is active or not before detection. ❖ Classical ML techniques still suffer from the lack of efficiency in detecting zero-day phishing attacks .	We used to determine the URL is active or not before detection.