LITERATURE SURVEY

Web phishing Detection - Phishing is a cyber attack to steal user's or organization data by replicating the original website to get the critical information.To prevent that web phising detection have been developed.

1.Farashazillah Yahya,Ryan Isaac W Mahibol,Chong Kim Ying,Magnus Bin Anai,Sidney Allister Frankie,Eric Ling Nin Wei and Rio Guntur Utomo**,"Detection of Phising Websites using Machine Learning Approaches",**2021 International Conference on Data Science and Its Applications (ICoDSA).

The Phishing websites can be detected by classifying them into legitimate or illegitimate websites using machine techniques.. The purpose is to review the existing techniques and implement experiments to detect whether a website is malicious or not.This paper implemented three supervising techniques which are Decision Tree , K-Nearest Neighbour (KNN), and Random Forest. These three algorithms are chosen because it provides a better understanding and is more suitable for the dataset.

2.Prajakta Patil,Rashmi Rane and Madhuri Bhalekar,"**Detecting spam and phishing mails using SVM and obfuscation URL detection algorithm",**2017 International Conference on Inventive Systems and Control (ICISC).

This paper aims to use fundamental visual features like page layouts and contents of a web page's appearance as the basis to detect page similarities.The standard way to specify page layouts is through the style sheet such that they develop an algorithm to detect similarities in key elements related to CSS.This paper proposed the SVM technique along with map-reduce paradigm to achieve a higher accuracy in detecting the spam email. By using map-reduce techniques they also try to overcome the two hurdles of the SVM.

3.Gaurav Varshney, Manoj Mishra and Pradeep K. Atrey, **"A phish detector using lightweight search features"**, Computers & Security, 2016.

Various web phishing detection solutions have been implemented which are quite heavy in terms of computational and communication requirements.It has been observed that search engine based solutions are the most lightweight and viable. This paper advances search engine based anti phishing research and presents the lightest possible phishing detection system, named the Lightweight Phish Detector (LPD).It runs on client browser for phishing detection.Exhaustive testing performed and a true positive rate obtained from phishtank URLs,

4.Antonio Hernández Dominguez and Walter Baluja García, **"Updated Analysis of Detection Methods for Phishing Attacks"**, Futuristic Trends in Network and Communication Technologies, vol.1395, pp.56, 2021.

This paper brings an updated study of the main existing mechanisms for the detection of phishing.Additionally,matching solutions for different services will be identified and the most effective solutions will be featured, with the aim of applying these approaches in future integrated solutions for the detection of phishing attacks.

5.Anggit Ferdita Nugraha and Luthfia Rahman, **"Meta-Algorithms for Improving Classification Performance in the Web-phishing Detection Process"**, 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp.271-275, 2019.

Research related to web phishing detection system has been carried out by many researchers, one of them using data mining techniques, but still uses a single classification algorithm. The meta-algorithm is proposed in this paper to support the improvement of classification performance for the development of various web phishing detection systems.

6.Yoga Pristyanto and Akhmad Dahlan, **"Hybrid Resampling for Imbalanced Class Handling on Web Phishing Classification Dataset"**, 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp.401-406, 2019.

The previous works which are related to web phishing,the researchers overlooked the imbalance class problem on the dataset.The majority of classification methods assume the nature of the class is balanced which will be declining the classification performance method.Therefore, the mechanism of imbalanced class handling is severely needed. In this study they used One Sided-Selection and Synthetic Minority Over-Sampling Technique to work together to handle the imbalanced class condition so that the accuracy and the g-mean score of the classification will be enhanced. Hence, the combination of OSS and SMOTE can be a plausible option to handle the imbalanced class problem on the web phishing classification either on binary class and multiclass datasets.

7.Athulya A.A and Praveen K,"**Towards the Detection of Phishing Attacks",**2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)

This paper discusses various phishing attacks, some of the latest phishing evasion techniques used by attackers and anti-phishing approaches. This review raises awareness of those phishing strategies and helps the user to practice phishing prevention. Here, a hybrid approach of phishing detection also described having fast response time and high accuracy.

8.Miyamoto D, Hazeyama H and Kadobayashi Y,**" An evaluation of machine learning-based methods for detection of phishing sites"** ,International Conference on Neural Information Processing pp. 539-546. Springer, Berlin, Heidelberg. (2008)

This paper presents the performance of machine learning-based methods for detection of phishing sites.It employs 9 machine learning techniques and combines these techniques with heuristics.It let's machine learning-based detection methods distinguish phishing sites from others. It analyzes the dataset composed of 1,500 phishing sites and 1,500 legitimate sites, which are classified by machine learning-based detection methods, and measures the performance.

9.K S Swarnalatha,K C Ramchandra,Kaushar Ansari,Love Ojha and Sanjok Subedi Sharma,"**Real-Time Threat Intelligence-Block Phising Attacks",**2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)

As of 2020, phishing is the most common attack performed by cyber criminals according to the FBI's Internet Crime Complaint Centre. First phishing Datasets are collected from phish tank and then legitimate websites are collected from University of New Brunswick and then dataset is preprocessed using wrapper and filter method so that it covers the dataset which gets missed, tampered and unstructured.

10.Salvi Siddhi Ravindra, Shah Juhi Sanjay, Shaikh Nausheenbanu Ahmed Gulzar and Khodke Pallavi, **"Phishing Website Detection Based on URL"**, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp.589, 2021.

This paper will introduce a possible solution to avoid phishing attacks by checking whether the provided URLs are phishing URLs or legitimate URLs. It is a Machine Learning based system especially Supervised learning where they have provided 2000 phishing and 2000 legitimate URL dataset. Random Forest Algorithm also have been taken into consideration due to its performance and accuracy. It considers the 9 features and detects whether the URL is safe or it is a phishing URL.