**LITERATURE SURVEY**

**Rao et al.**proposed a novel classification approach that use heuristic based feature extraction approach. In this, they have classified extracted features into three categories such as URL Obfuscation features, Third-Party-based features, Hyperlink-based features. Moreover, proposed technique gives 99.55% accuracy. Drawback of this is that as this model uses third-party features, classification of website dependent on speed of third-party services. Also this model is purely depends on the quality and quantity of the training set and Broken links feature extraction has a limitation of more execution time for the websites with more number of links.

**Priyanka et al.**proposed novel approach by combining two or more algorithms. In this paper ,author has implemented two algorithm Adaline and Backpropion along with SVM for getting good detection rate and classification purpose.

**Nepali and Wang** proposed a novel approach to detect malicious URLs using only visible features from social networks.

**Kuyama et al** proposed a method for identifying the Command and Control server (C&C server) by using supervised learning and features points obtained from WHOIS and DNS information. They evaluated domain names and email addresses from the WHOIS as input values for machine learning.

**Sadeh et al.**proposed a system called PILFER for classifying phishing URLs. They extracted a set of ten features that are specifically designed to highlight deceptive methods used to fool users. The data set consists of approximately 860 phishing e-mails and 6950 non phishing emails. They used a Support Vector Machine (SVM) as a classifier in the implementation. They trained and tested the classifier using 10-fold cross validation and obtained 92 percent accuracy.

**Ma et al.** considered the URL classification problem as a binary classification problem and built a URL classification system that processes a live feed of labeled URLs. It also collects URL features in real time from a large Web mail provider. They used both lexical and host-based features. From the gathered features and labels, they were able to train an online classifier using a Confidence Weighted (CW) algorithm.

**Parkait et al.** provide a comprehensive literature review after analyzing 358 research papers in the area of phishing counter measures and their effectiveness. They classified antiphishing approaches into eight groups and highlighted advanced anti-phishing methods.

**Abdelhamid et al.** built a system for detecting phishing URLs called Multi-label Classifier based on Associative Classification (MCAC). They used sixteen features and classified URLs into three classes: phishing, legitimate, and suspicious. The MCAC is a rule-based algorithm where multiple label rules are extracted from the phishing data set.

**Patil** provided a brief overview of various forms of web-page attacks in their survey on malicious webpages detection techniques.

**Hadi et al.**used the Fast-Associative Classification Algorithm (FACA) for classifying phishing URLs. FACA works by discovering all frequent rule item sets and building a model for classification. They investigated a data set consisting of 11,055 websites with two classes, legitimate and phishing. The data set contained thirty features. They used the minimum support and the minimum confidence threshold values as two percent and fifty percent, respectively.

**Chunlin et al.** proposed approach that primarily focus on character frequency features. In this they have combined statistical analysis of URL with machine learning technique to get result that is more accurate for classification of malicious URLs. Also they have compared six machine-learning algorithms to verify the effectiveness of proposed algorithm which gives 99.7% precision with false positive rate less than 0.4%.

**Bhagyashree et al**.proposed a feature based approach to classify URLs as phishing and nonphishing. Various features this approach uses are lexical features, WHOIS features, Page Rank and Alexa rank and Phish Tank-based features for disguising phishing and non-phishing website. In this paper, web-mining classification is used.

**M. Amaad et al.**presented a hybrid model for classification of phishing website. In this paper, proposed model carried out in two phase. In phase 1,they individually perform classification techniques, and select the best three models based on high accuracy and other performance criteria. While in phase 2, they further combined each individual model with best three model and makes hybrid model that gives better accuracy than individual model. They achieved 97.75% accuracy on testing dataset. There is limitation of this model that it requires more time to build hybrid model.

**Hossein et al.** developed an open-source framework known as “Fresh-Phish”. For phishing websites, machine-learning data can be created using this framework. In this, they have used reduced features set and using python for building query .They build a large labelled dataset and analyses several machine- learning classifiers against this dataset .Analysis of this gives very good accuracy using machine-learning classifiers. These analyses how long time it takes to train the model.

**Luong et al.** proposed new technique to detect phishing website. In proposed method, Author used six heuristics that are primary domain, sub domain, path domain, page rank, and Alexa rank, Alexa reputation whose weight and values are evaluated. This approach gives 97 % accuracy but still improvement can be done by enhancing more heuristics.

**Ahmad et al.** proposed three new features to improve accuracy rate for phishing website detection. In this paper, Author used both type of features as commonly known and new features for classification of phishing and non-phishing site. At the end author has concluded this work can be enhanced by using this novel features with decision tree machine learning classifiers.

**Mohammad et al**.proposed model that automatically extracts important features for phishing website detection without requiring any human intervention. Author has concluded in this paper that the process of extracting feature by their tool is much faster and reliable than any manual extraction.

REFERENCES

1. Phishing definition, https://en.wikipedia.org/wiki/Phishing
2. APWG Report 1, http://docs.apwg.org/reports/apwg\_trends\_repo rt\_q2\_2018.pdf
3. APWG report 2, http://docs.apwg.org/reports/apwg\_trends\_repo rt\_q1\_2018.pdf
4. J. Han and M. Kamber, Data Mining Concepts and Techniques, Elsevier, 2006.

[5] Phishing dataset, https://www.phishtank.com/developer\_info.ph p

[5] Routhu Srinivasa Rao1 , Alwyn Roshan Pais : Detection of phishing websites using an efficient feature-based machine learning framework :In Springer 2018.

[6] D. R. Patil and J. Patil, J., “Survey on malicious web pages detection techniques”, International Journal of u-and e-Service, Science and Technology, vol. 8, no. 5, pp. 195–206, 2015.

[7] W. Hadi, F. Aburrub, and S, Alhawari, “A new fast associative classification algorithm for detecting phishing websites”, Applied Soft Computing vol. 48, pp 729-734, 2016.

[8] R. K. Nepali and Y. Wang, Y., “You look suspicious!! Leveraging visible attributes to classify malicious short urls on twitter”, 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, pp. 2648–2655, 2016.

[9] M. Kuyama, Y. Kakizaki, and R. Sasaki, “Method for detecting a malicious domain by using whois and dns features”, The Third International Conference on Digital Security and Forensics (DigitalSec2016), p. 74, 2016.

[10] S. Purkait, “Phishing counter measures and their effectiveness–literature review”, Information Management & Computer Security, vol. 20, no. 5, pp. 382–420, 2012.

[11] N. Abdelhamid, A. Ayesh, F. Thabtah, “Phishing Detection based Associative Classification”, Data Mining. Expert Systems with Applications (ESWA), vol. 41, pp 5948-5959, 2014.

[12]<https://www.researchgate.net/profile/Er-Purvi-Pujara/publication/331198983_Phishing_Website_Detection_using_Machine_Learning_A_Review/links/5c6bd4ae4585156b5706e727/Phishing-Website-Detection-using-Machine-Learning-A-Review.pdf>

[13]https://www.researchgate.net/profile/Akila-D/publication/337049054\_Phishing\_Websites\_Detection\_Using\_Machine\_Learning/links/5dc258bd92851c8180304c07/Phishing-Websites-Detection-Using-Machine-Learning.pdf