iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40828

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Detection of URL Based Phishing Websites Using Machine Learning in Django Framework

Khushboo Kumari[1], Feon Jaison[2]
[2]*Assistant Professor,* [1]*Department of Master of Computer Applications, School of CS & IT JAIN (Deemed-to-be-University)*

Abstract: ***In this modern world, Phishing website detection is one of the most critical tasks in the world. In the recent times, a lot of people have suffered phishing attack due to phishing website. Machine Learning plays an important role in prediction of phishing website in the network. The proposed method predicts the URL based phishing websites based on features and also gives maximum accuracy to predict the result. This method uses uniform resource locator (URL) features to detect. It identified features that phishing site URLs contain. The proposed method takes those features for phishing detection. Security of the phishing detection website is also a major concern which is solved by providing administration who can manage the phishing detection website.***

Keywords: Phishing site, Machine learning, Legitimate, Prediction

I. INTRODUCTION

The work proposed in this model focus on URL based phishing website in machine learning prediction. These phishing websites are made to look the appearance of an original organization website. The attacker force user to fill up the personal information by giving alarming messages or validate account messages or need urgent update etc so that they fill up the required information which can be used by them to misuse it. They make the situation such that the user is not left with any other option but to visit their fake website. Machine learning has been widely used in many areas to create automated solutions. The phishing attacks can be carried out in many ways such as email, website, malware and voice. Similarly, the samples which are labelled as legitimate will be detected as legitimate URL. The dataset to be used for machine learning must consist these features. There so many machine learning algorithms and each algorithm has its own working mechanism

The algorithm used are logistic regression, super vector machine to predict the phishing site. For security of the phishing detection website the administration and password are used to protect the detection website.

II. LITERATURE REVIEW

Ashit Kumar Dutta et.al [3] introduced a Detection and Prevention Of phishing Website using ML using Machine Learning. The author proposed a URL detection technique based on machine learning approaches. A recurrent neural network method is employed to detect phishing URL. We adopt the Detection procedure and by analysing the results, it is predicted to provide security to URL detection website.

Jain A.K et.al [4] proposed a URL-based anti-phishing machine learning method. They have taken 14 features of the URL to detect the website as a malicious or legitimate to test the efficiency of their method. More than 33,000 phishing and valid URLs in Support Vector Machine (SVM) and Naïve Bayes (NB) classifiers were used to train the proposed system. The phishing detection method focused on the learning process. They extracted 14 different features, which make phishing websites different from legitimate websites. The outcome of their experiment reached over 90% of precision when websites with SVM Classification are detected.

Rishikesh Mahajan et.al [1] proposed paper that deals with machine learning technology for detection of phishing URLs by extracting and analysing various features of legitimate and phishing URLs. Decision Tree, random forest and Support vector machine algorithms are used to detect phishing websites. To improve the extracting, we used some more algorithm.

Anmol Chhetri et.al [3] are discussing and listing a few of the artificial intelligence models, that will help us to detect these phishing websites so that in the future these data and techniques can be used in machine learning to make our system better and efficient. The

Juan Chen et.al [2] propose a new end-host based anti-phishing algorithm, which we call Link Guard, by utilizing the generic characteristics of the hyperlinks in phishing attacks.

Muhammed Baykara et.al [2] proposed Detection of phishing attacks proposed phishing and spam mails are detected by examining mail contents. Classification of spam words added to the database by Bayesian algorithm is provided.

©IJRASET: All Rights are Reserved

Athulya A A et.al [2] proposed towards the detection of phishing attacks proposed as applying every technique in a single phase is time-consuming, will pick one technique at a given time based on condition. URL entered will be checked by the Blacklist method or whitelist method or search engine-based technique.

By analysing all the papers, we provided with the security feature in the detection website as an administration and password for admin to login and able to manage the number of users and able to read the feedback from the user if there is any feedback relate to login or website.

The user can login with username and password and able to check phishing website and provide feedback related to phishing website.

The prediction of phishing website has found that system provides us with 95 % of accuracy for Multinomial NB and 97 % of accuracy for Logical Regression. The prediction is done between good and bad website and training accuracy is 97% and testing accuracy is 95%.

Author	ML Models	Accuracy Rate %	Contribution	Limitation
Rishikesh Mahajan et.al [1]	Decision tree	96.71	Dataset is divided into training set and testing set in 50:50, 70:30 and 90:10 ratios	hybrid technology does not implement to detect phishing websites more accurately.
	Random tree	96.72		
	Support vector machine	96.40		
Ashit Kumar Dutta et.al [3]	LSTM technique	96.4	considered to involve automatic categorization of websites into a predetermined set of class values based on several features and the class variable.	Does not generate an outcome for a larger network and protect the privacy of an individual.
Jain A.K. et.al [4]	Naïve Bayes	95	Employed both NB and SVM algorithms to identify the malicious websites.	Both SVM and NB are slow learners and does not store the previous results in the memory. Thus, the efficiency of the URL detector may be reduced.
	Support Vector Machine	90		
Hung Le et al., [5]	Naïve Bayes	96.92	Developed a crawler to extract URLs from data repositories. Applied lexical features approach to identify the phishing websites.	The performance evaluation was based on crawler-based dataset. Thus, there is no assurance for the effectiveness of the URL detector with real time URLs.
	Support vector machine	95.35		
	Random tree	95		
Kumar J. et al., [6]	Random tree	95.46	Proposed a URL detector based on blacklisted dataset. Also, a lexical feature approach was employed to classify malicious and legitimate websites.	Authors employed an older dataset which can reduce the performance of the detector with real—time URLs.
	Support Vector Machine	95.22		

III. EXISTING METHODOLOGIES

Three machine learning classification models which is used to predict phishing URL are Decision Tree, Random Forest and Support vector machine.

A. Decision Tree Algorithm

One of the most widely used algorithm in machine learning technology. Decision tree algorithm is easy to understand and easy to implement. Decision tree do its work by choosing best splitter from the available attributes for classification which can be considered as a root of the tree. Algorithm continues to build tree until it finds the last leaf node. Decision tree creates training model which is used to predict target value or class in a tree representation each internal nodes of the tree belongs to attribute and each leaf node of the tree belongs to class label.

B. Random Forest Algorithm

Random forest algorithm is one of the most powerful algorithms in machine learning technology and it is based on concept of decision tree algorithm. This algorithm creates the forest with number of decision trees. High number of trees gives high detection accuracy. Creation of trees are based on bootstrap method. In bootstrap method features and samples of dataset are randomly selected with replacement to build a single tree. In randomly selected features, random forest algorithm will choose best splitter for the classification and like decision tree algorithm; Random Forest algorithm information gain methods to find the best splitter. This process will get continue until random forest creates n number of trees. Each tree in forest predicts the target value and then algorithm will be calculating the votes for each predicted target. Finally random forest algorithm considers high voted predicted target as a final prediction.

C. Support Vector Machine Algorithm

Support vector machine is another powerful algorithm in machine learning technology. In support vector machine algorithm, each data item is plotted as a point in n-dimensional space and support vector machine algorithm built separating line for classification of two classes, this separating line is well known as hyperplane. Support vector machine seeks for the closest points called as support vectors and once it finds the closest point it draws a line connecting to them. This machine then constructs separating line which bisects and perpendicular to the connecting line. In order to classify data perfectly the margin should be maximum. Here the margin is a distance between hyperplane and support vectors. In real scenario it is not possible to separate complex and non-linear data, to solve this problem support vector machine uses kernel trick which transforms lower dimensional space to higher dimensional space.

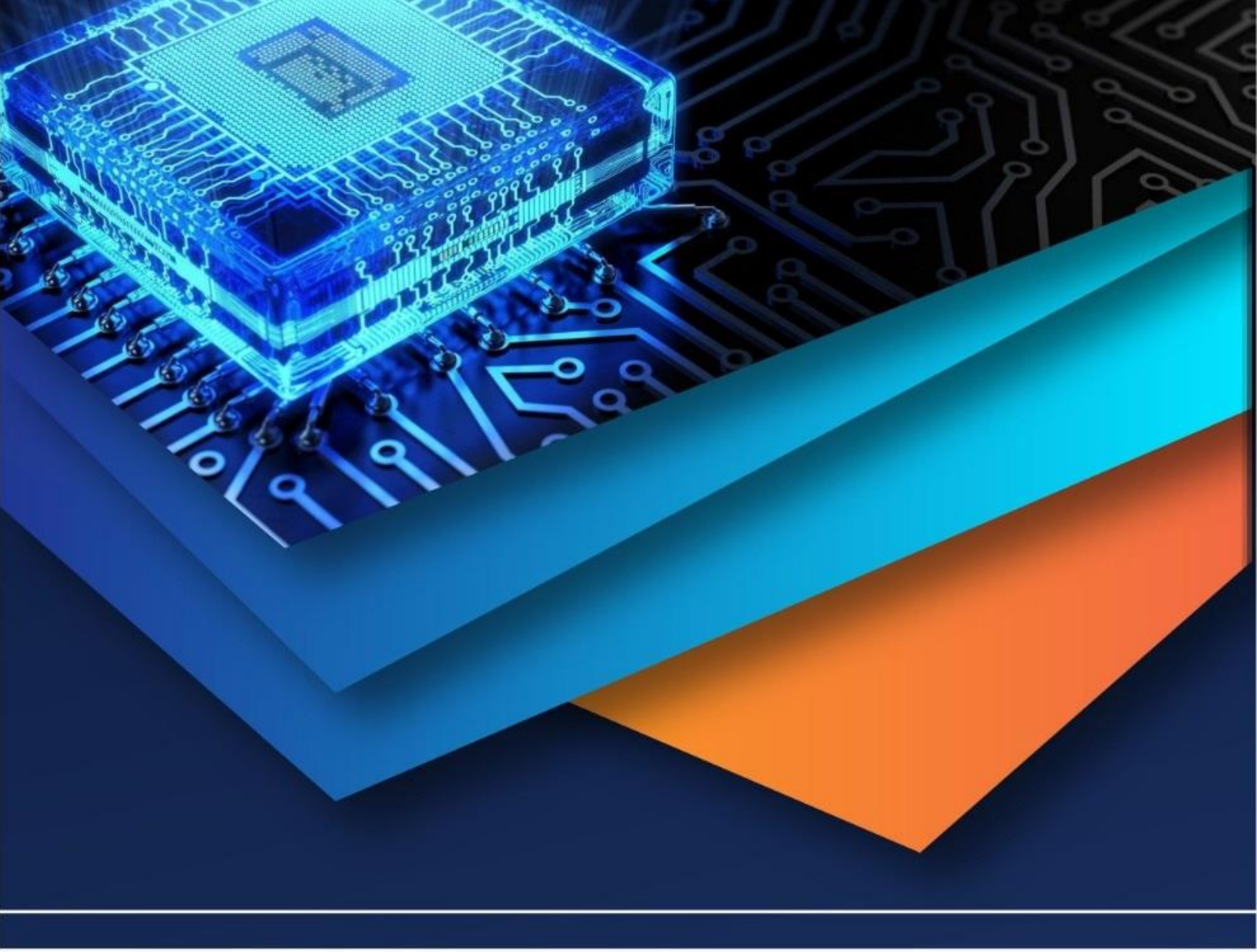
IV. CONCLUSION

It is found that phishing attacks is a critical task and it is important for us to get a method to detect it. This issue can be easily solved by using any of the machine learning algorithm with the classifiers. Classifiers which give good prediction rate of the phishing website, but after our survey that it will be better to use a hybrid approach such as NB and Multinomial NB for the prediction and more improve the accuracy prediction rate of phishing websites. Existing methodologies gives less accuracy so we proposed a new phishing method with security for the admin for URL based features and machine learning algorithms. The Security for the website which is used to detect phishing website is done by admin and user part in Django framework.

REFERENCES

[1] Rishikesh Mahajan, Irfan Siddavatam, “Phishing website detection using machine learning” International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 23, October 2018
[2] Hodžić, A., Kevrić, J., & Karadag, A. (2016). Comparison of machine learning techniques in phishing website classification. In International Conference on Economic and Social Studies (ICESoS'16) (pp. 249-256).
[3] Ashit Kumar Dutta,” Detecting phishing Website using machine learning technique” Published online 2021 Oct 11. Doi: 10.1371/journal.pone.0258361 PMCID: PMC8504731 PMID: 34634081
[4] Jain A.K., Gupta B.B. “PHISH-SAFE: URL Features-Based Phishing Detection System Using Machine Learning”, Cyber Security. Advances in Intelligent Systems and Computing, vol. 729, 2018, Doi: 10.1007/978-981-10-8536-9_44
[5] Hung Le, Quang Pham, Doyen Sahoo, and Steven C.H. Hoi, “URL Net: Learning a URL Representation with Deep Learning for Malicious URL Detection”, Conference’17, Washington, DC, USA, arXiv:1802.03162, July 2017.
[6] J. Kumar, A. Santhanavijayan, B. Janet, B. Rajendran and B. S. Bindhumadhava, “Phishing Website Classification and Detection Using Machine Learning,” 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1–6, 10.1109/ICCCI48352.2020.9104161.
[7] Wu CY, Kuo CC, Yang CS,” A phishing detection system based on machine learning” In: 2019 International Conference on Intelligent Computing and its Emerging Applications (ICEA), pp 28–32, 2019.
[8] Rao RS, Pais AR. Jail-Phish: An improved search engine-based phishing detection system. Computers & Security. 2019. Jun 1;83:246–67. [Google Scholar]
[9] Aljofey A, Jiang Q, Qu Q, Huang M, Niyigena JP. An effective phishing detection model based on character level convolutional neural network from URL. Electronics. 2020. Sep;9(9):1514. [Google Scholar]
[10] Praisy Evangelin A1, Jeenath Laila N: PRIVACY PROTECTION OF USER BROWSING DETAILS AND UNSAFE URL DETECTION. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 03 | Mar 2021

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)