Journal of Nonlinear Analysis and Optimization
Vol. 15, Issue. 1 : 2024
ISSN : **1906-9685**

Journal of Nonlinear Analysis and Optimization : Theory & Applications

Editors-in-Chief :

IDENTIFYING FRAUDULENT WEBSITES WITH FEATURE-BASED MACHINE LEARNING

#1 MR.K.MAHANTHI, #2 A.SAI DEEPTHI, #3 G.SAI VARA PRASAD, #4 B.YASHASWINI, #5 K.PAVAN SUBBARAO
#1Assistant professor in Department of IT, DVR & Dr.HS MIC College of Technology,Kanchikacherla
#2#3#4#5 B.Tech with Specialization of Information Technology , DVR & Dr.HS MIC College of Technology,Kanchikacherla-521180

ABSTRACT_ One of the most frequent and harmful types of cyberattacks is phishing. The intention behind these attacks is to pilfer the data that people and businesses use to carry out transactions. These phishing websites include a variety of clues in their text and browser-based data. The fraudsters send information from a phoney website or email to the intended recipient posing as a bank, organisation, or other trustworthy source that handles trustworthy transactions.The aim of this research is to use a Random Forest classifier to perform Extreme Learning Machine (ELM) based categorization for 30 characteristics, including Internet worm Websites Data in the UC Irvine Machine Learning Repository database. The proposed study is predicated on a phishing URL-based dataset that was taken from a well-known dataset repository. This dataset comprises vectorized phishing and legal URL features that were gathered from over 11,000 website datasets. To increase accuracy, we extend the approach using the CatBoost classifier in the suggested method.

1.INTRODUCTION

Web use has turned into a fundamental piece of our day to day exercises because of quickly developing innovation. Because of this fast development of innovation and concentrated utilization of advanced frameworks, information security of these frameworks has acquired extraordinary significance. The essential target of keeping up with security in data advancements is to guarantee that vital safety measures are taken against dangers and perils liable to be looked by clients during the utilization of these innovations [1]. Phishing is characterized as mimicking solid sites to acquire the restrictive data went into sites consistently for different purposes, for example, usernames, passwords and citizenship numbers. Web worm sites contain different clues among their items and internet

browser-based data [2-4]. Individual(s) committing the misrepresentation sends the phoney site or email data to the objective location as though it comes from an association, bank or whatever other solid source that performs dependable exchanges.

Items in the site or the email incorporate solicitations planning to bait the people to enter or refresh their own data or to change their passwords as well as connections to sites that seem to be precise duplicates of the sites of the associations concerned . Web worm Sites Highlights Many articles have been distributed about how to anticipate the Web worm sites by utilizing computerized reasoning strategies. We analyzed Web worm sites and extricated highlights of these sites Rules with respect to the removed highlights of this information base are given underneath. In the primary area we characterized rules and we gave conditions of web highlights. We really want these conditions to make sense of Web worm assaults characaterization.

2.LITERATURE SURVEY

Intelligent rule based Internet worm websites classification

Internet worm is described as the art of echoing a website of a creditable firm intending to grab user's private information such as usernames, passwords and social security number. Internet worm websites comprise a variety of cues within its content-parts as well as the browser-based security indicators provided along with the website. Several solutions have been proposed to tackle Internet worm . Nevertheless, there is no single magic bullet that can solve this threat radically. One of the promising techniques that can be employed in predicting Internet worm attacks is based on data mining, particularly the `induction of classification rules' since anti-Internet worm solutions aim to predict the website class accurately and that exactly matches the data mining classification technique goals. In this study, the authors shed light on the important features that distinguish Internet worm websites from legitimate ones and assess how good rule-based data mining classification techniques are in predicting Internet worm websites and which classification technique is proven to be more reliable.

Predicting Phishing websites based on self-structuring neural network

we proposed an intelligent model for predicting Internet worm attacks based on artificial neural network particularly self-structuring neural networks. Internet worm is a continuous problem where features significant in determining the type of web pages are constantly changing. Thus, we need to constantly improve the network structure in order to cope with these changes. Our model solves this problem by automating the process of structuring the network and shows high acceptance for noisy data, fault tolerance and high prediction accuracy. Several experiments were conducted in our research, and the number of epochs differs in each experiment. From the results, we find that all produced structures have high generalization ability.

3.PROPOSED SYSTEM

The following 30 features were retrieved from webpages in the UC Irvine Machine Learning Repository and classified using an Extreme Learning Machine (ELM) in this study. To improve accuracy in our investigation, we are utilising the CATBOOST classifier. The following are the procedural stages for resolving the given classification problem:

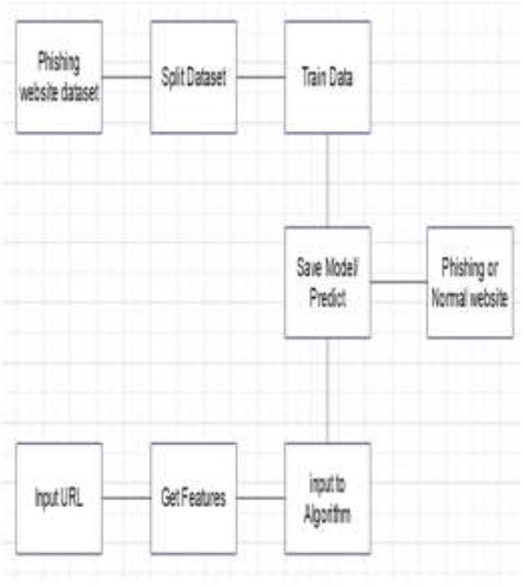
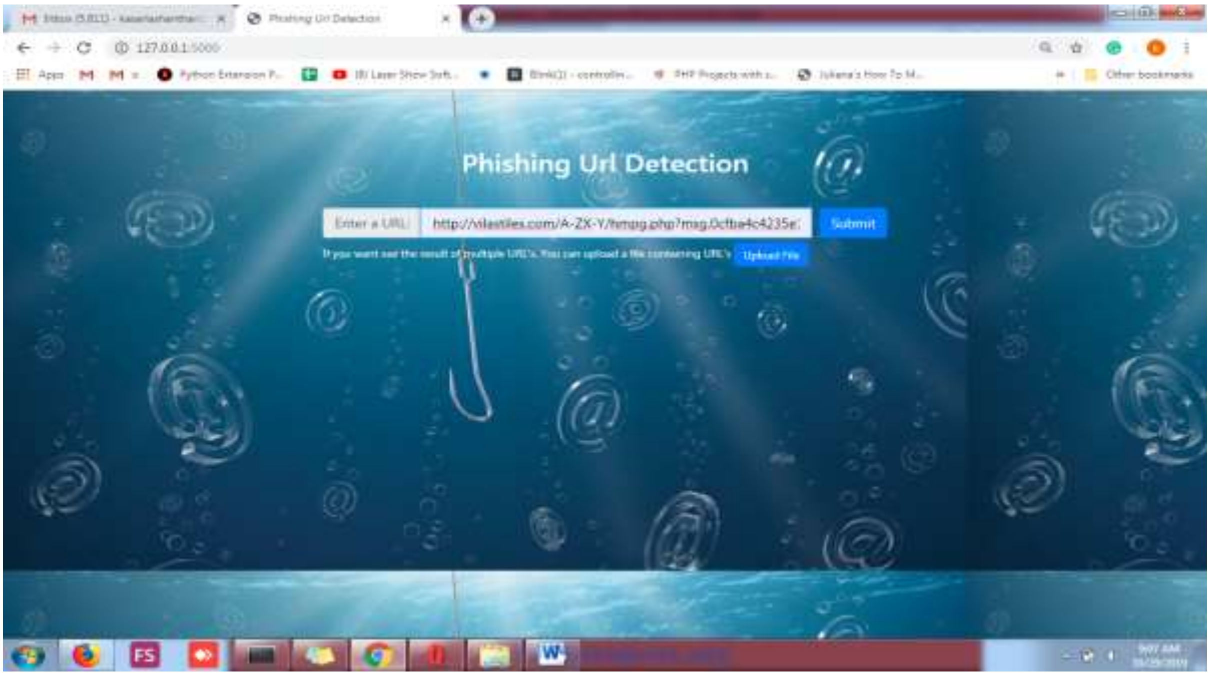


Fig 1:Architecture

3.1 IMPLEMENTATION

- **User module:**
  Can Using this module user will have a user web interface where he can enter any website URL and check if that given link is legitimate or internet worm website.
- **Feature Training Module:**
  In this module phishing website URL dataset is taken and for each website 30 features are extracted from the data set and for each feature output will be either -1,0,1 and these details are stored in csv file format. This data is used as training data for detecting internet worm website.
- **Phishing Website Detection Module:**
  When user enters URL link for given link 30 features are extracted using many libraries and output of -1,0,1 array list is sent to check with training data and predict phishing website using machine learning algorithm.

4.RESULTS AND DISCUSSION

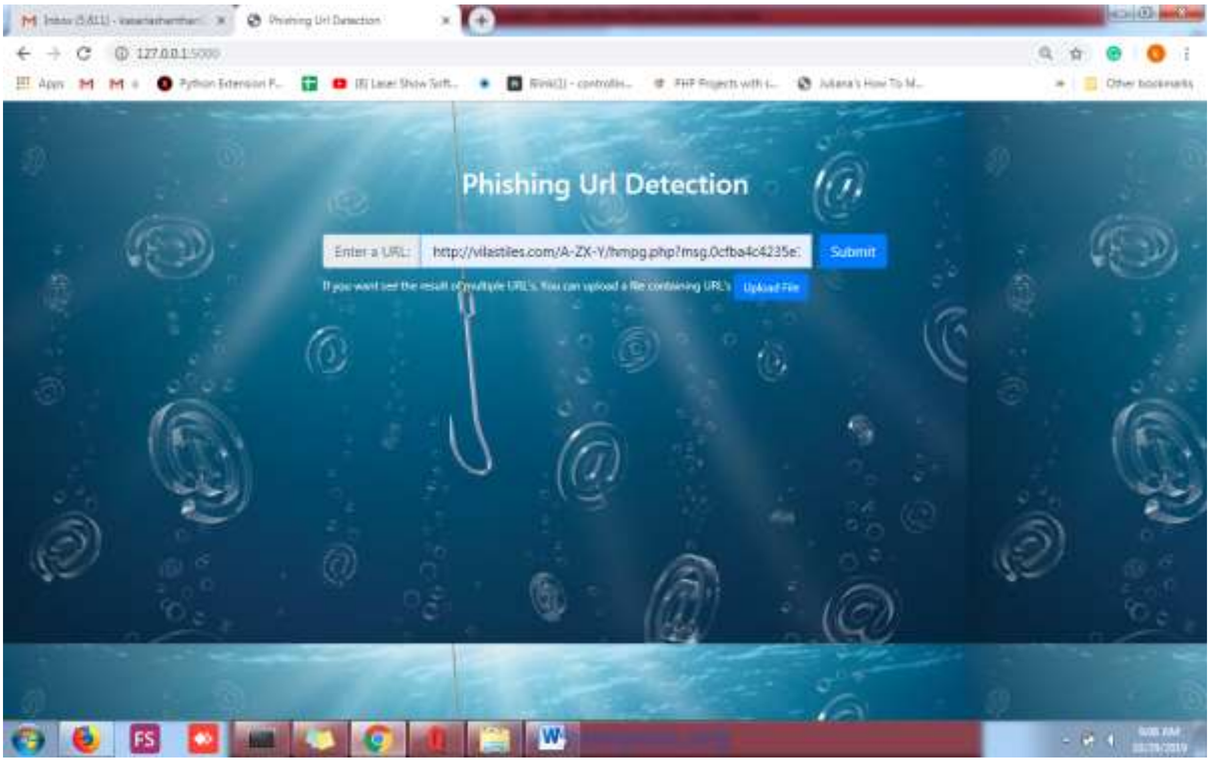


Phishing Url Detection

Enter a URL: http://vilastiles.com/A-ZX-Y/hmpg.php?msg.0cfba4c4235e

If you want see the result of multiple URL's. You can upload a file containing URL's

After entering click on submit

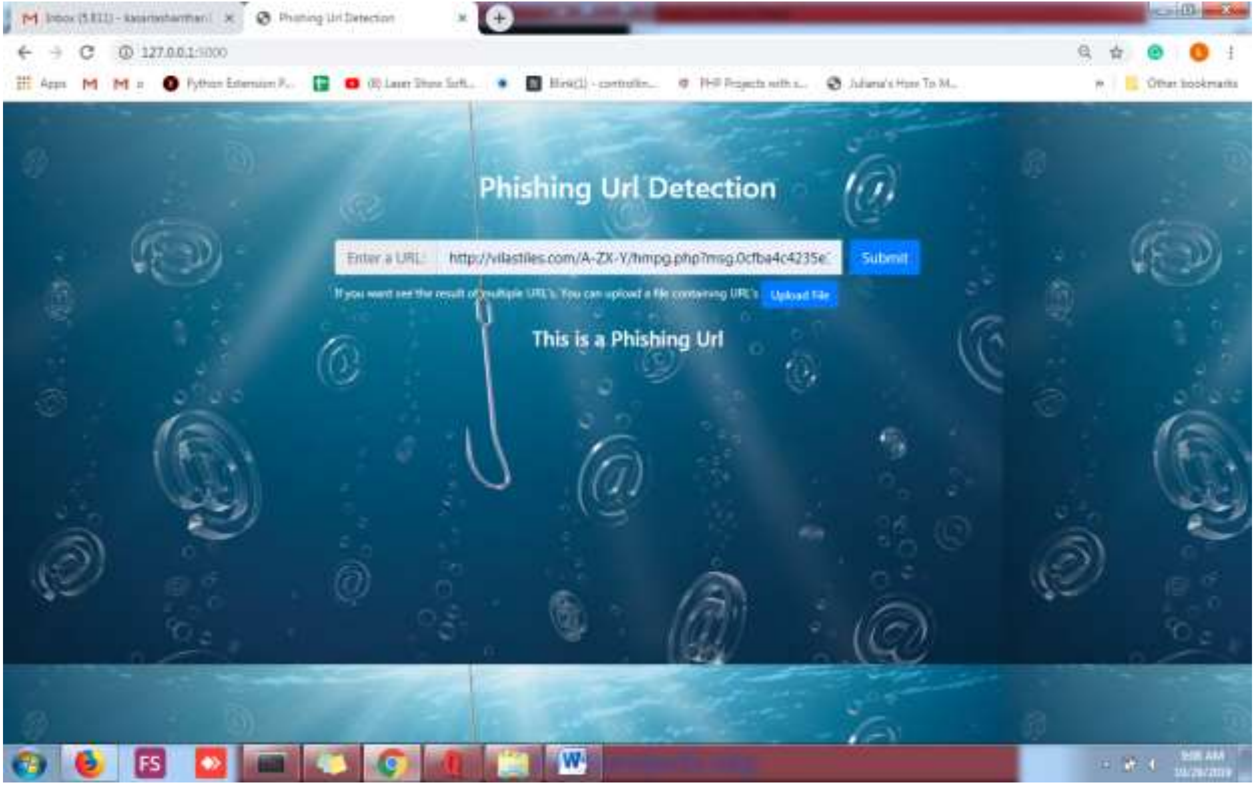


Phishing Url Detection

Enter a URL: http://vilastiles.com/A-ZX-Y/hmpg.php?msg.0cfba4c4235e

If you want see the result of multiple URL's. You can upload a file containing URL's

Internet worm website result



Phishing Url Detection

Enter a URL: http://vilastiles.com/A-ZX-Y/hmpg.php?msg.0cfba4c4235e

If you want see the result of multiple URL's. You can upload a file containing URL's

This is a Phishing Url

5.CONCLUSION

In order to classify Internet worm assaults, we created a classification model and outlined characteristics of Internet worm attacks in this study. This method includes a categorization part and feature extraction from websites. Our well-defined Internet worm feature extraction rules have been applied to the feature extraction process in order to get features. SVM, NB, and ELM were utilised in the classification of these features. Six distinct activation functions were employed in the ELM, and it obtained the greatest accuracy score.

REFERENCES

[1] G. Canbek and "A Review on Information, Information Security and Security Processes," Politek. Derg., vol. 9, no. 3, pp. 165–174, 2006.

[2] L. McCluskey, F. Thabtah, and R. M. Mohammad, "Intelligent rulebased Internet worm websites classification," IET Inf. Secur., vol. 8, no. 3, pp. 153–160, 2014.

[3] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting Internet worm websites based on self-structuring neural network," Neural Comput. Appl., vol. 25, no. 2, pp. 443–458, 2014.

[4] R. M. Mohammad, F. Thabtah, and L. McCluskey, "An assessment of features related to Internet worm websites using an automated technique," Internet Technol. …, pp. 492–497, 2012.

[5] W. Hadi, F. Aburub, and S. Alhawari, "A new fast associative classification algorithm for detecting Internet worm websites," Appl. Soft Comput. J., vol. 48, pp. 729–734, 2016.

[6] N. Abdelhamid, "Multi-label rules for Internet worm classification," Appl. Comput. Informatics, vol. 11, no. 1, pp. 29–46, 2015.

[7] N. Sanglerdsinlapachai and A. Rungsawang, "Using domain top-page similarity feature in machine learning-based web Internet worm detection," in 3rd International Conference on Knowledge Discovery and DataMining, WKDD 2010, 2010, pp. 187–190.]

1.1.1 Author's Profiles

#1:-MR.K.MAHANTHI working as Assistant Professor in Department of IT in DVR & Dr,HS MIC College of Technology,Kanchikacherla-521180

#2:-A.SAI DEEPTHI(20HA1A1232) B.Tech with Specialization of
Information Technology in DVR & Dr.HS MIC College of Technology,Kanchikacherla521180

#3:- G.SAI VARA PRASAD(20H71A1235) B.Tech with Specialization of
Information Technology in DVR & Dr.HS MIC College of Technology,Kanchikacherla-521180

#4:- B.YASHASWINI(20H71A1264) B.Tech with Specialization of
Information Technology in DVR & Dr.HS MIC College of Technology,Kanchikacherla521180

#5:- K.PAVAN SUBBARAO(20H71A1222) B.Tech with Specialization of
Information Technology in DVR & Dr.HS MIC College of
Technology,Kanchikacherla-521180