International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 03 | Mar 2021 www.irjet.net p-ISSN: 2395-0072

Phishing Website Detection based on Multidimensional Features Driven by Deep Learning

Buddhapuri Aneerudh[1], Bhumireddy Suneel Kumar Reddy[2], Dr.R.Radhika[3]

[1]*Student, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India*
[2]*Student, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India*
[3]*Assistant Professor, Dept. Of CSE, SCSVMV (Deemed to be University), Kanchipuram, TamilNadu, India*

Abstract – *Web Phishing appeals to the user to connect with the fake site. The main goal of this attack is to rob the user of sensitive information. The intruder builds websites similar to those that look like the original website. It allows attackers to access confidential information such as username, password, details of credit cards etc. This paper aims to review many of the phishing detection strategies recently suggested for the website. This will also provide a high-level description of various forms of phishing detection techniques. Here proposed a multidimensional element phishing recognition approach dependent on a quick discovery method by using deep learning (MFPD). In the initial step, character succession highlights of the given URL are separated and utilized for snappy characterization by profound learning, and this progression doesn't need outsider help or any earlier information about phishing. In the subsequent advance, we consolidate URL measurable highlights, website page code highlights, site page content highlights and the brisk characterization consequence of profound learning into multidimensional highlights. The methodology can diminish the identification time for setting an edge. Testing on a dataset containing a huge number of phishing URLs and genuine URLs, the exactness arrives at 98.99%, and the bogus positive rate is just 0.59%. By sensibly changing the limit, the test results show that the discovery effectiveness can be improved.*

***Key Words*: CNN-LSTM, CNN-BiLSTM and MFPD**

1. INTRODUCTION

The phishing website detection based on machine learning is a hotspot of current phishing website detection research. The results of machine learning methods usually depend on the quality of the extracted features. The focus of current research is on how to extract and select more effective features before processing them.

1.1 Scope of the project

Phishing website recognition framework gives solid security system to distinguish and forestall phishing areas from arriving at client. This venture presents a basic and compact way to deal with identify parodied pages and fathom security vulnerabilities utilizing Machine Learning. It very well may be effectively worked by anybody since all the significant errands are going on in the backend.

1.2 Modules Description

- Data Acquisition: Upload the URL data from the local host
- Data Preprocessing: In this module, we will perform label encoding, convert the text data into token counts and quantify a word in documents, we generally compute a weight to each word which signifies the importance of the word in the document and corpus.
- Splitting: In this module we will split the data into train and test data.
- Modelling: in this module, we will apply the CNN-LSTM and CNN-BiLSTM on URL text and we will apply the machine learning algorithms on the features of URL.
- Comparison: Visualize the varies accuracy of modeling
- Prediction: Url phishing detection on the new site

2. SOFTWARE REQUIREMENTS

The software tools that needs to be installed and used are listed here

- Python idle 3.7 version (or)
- Anaconda 3.7 (or)
- Jupiter (or)
- Google colab

2.2 Structure of Project

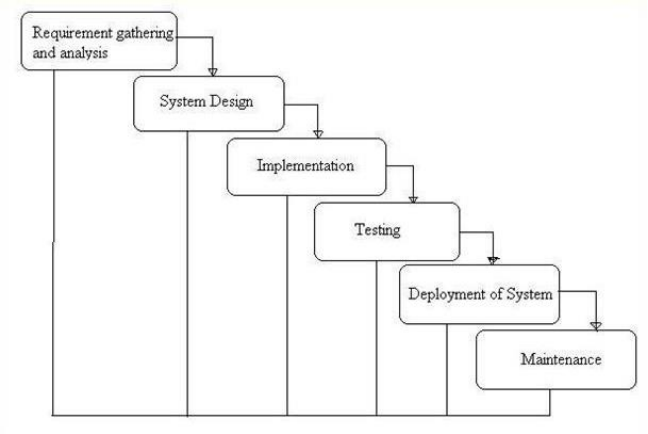


Chart -1: Project SDLC

- Project Requisites Accumulating and Analysis
- Application System Design
- Practical Implementation
- Manual Testing of My Application
- Application Deployment of System
- Maintenance of the Project

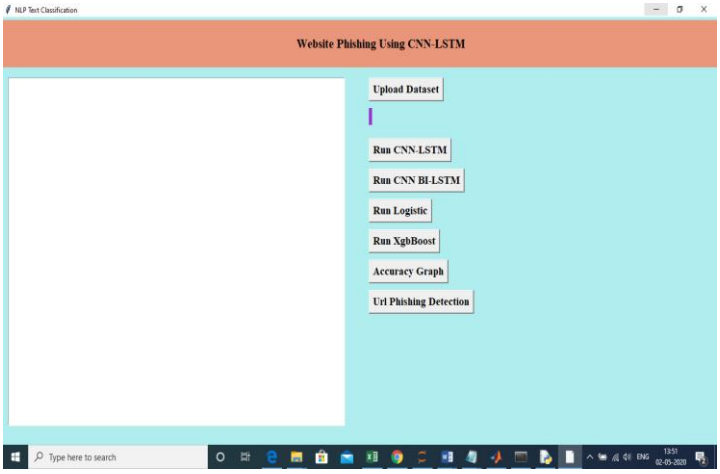
3. System Design

In System Design has divided into three types like GUI Designing, UML Designing with avails in development of project in facile way with different actor and its utilizer case by utilizer case diagram, flow of the project utilizing sequence, Class diagram gives information about different class in the project with methods that have to be utilized in the project if comes to our project our UML Will utilizable in this way The third and post import for the project in system design is Data base design where we endeavor to design data base predicated on the number of modules in our project

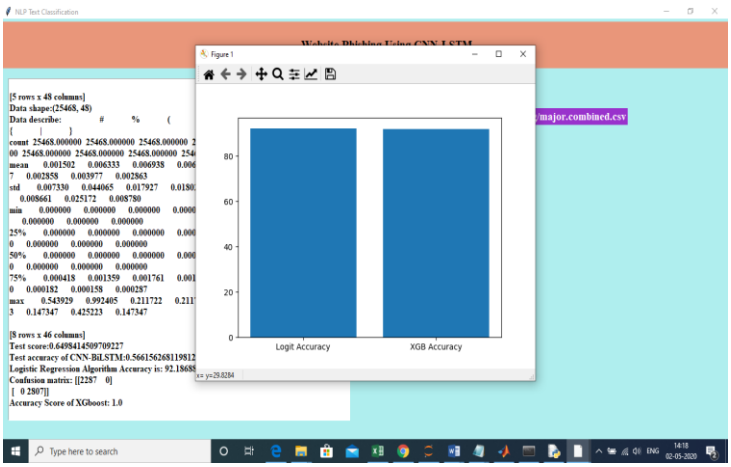
3.1 Implementation

The Implementation is Phase where we endeavor to give the practical output of the work done in designing stage and most of Coding in Business logic lay coms into action in this stage its main and crucial part of the project

4. Results



Upload and run



CONCLUSION

It is well known that a good phishing website detection approach should have good real-time performance while ensuring good accuracy and a low false positive rate. Our proposed MFPD approach is consistent with this idea. Under the control of a dynamic category decision algorithm, the URL character sequence without phishing prior knowledge ensures the detection speed, and the multidimensional feature detection ensures the detection accuracy. We conduct a series of experiments on a dataset containing millions of phishing and legitimate URLs. From the results, we find that the MFPD approach is effective with high accuracy, low false positive rate and high detection speed. A future development of our approach will consider applying deep learning to feature extraction of webpage code and webpage text. In addition, we plan to implement our approach into a plugin for embedding in a Web browser.

REFERENCES

[1] (2018). Phishing Attack Trends Re-Port-1Q. Accessed: May 5, 2018.

[Online]. Available: https://apwg.org/resources/apwg-reports/

[2] (2017). Kaspersky Security Bulletin: Overall Statisticals For. Accessed: Jul 12 2018. [Online]. Available: https://securelist.com/ksb-overallstatistics-2017/83453/

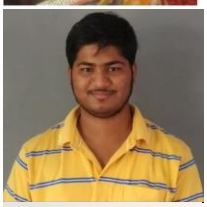
[3] A.Y. Ahmad, M. Selvakumar, A. Mohammed, and A.-S. Samer, ``TrustQR:

A new technique for the detection of phishing attacks on QR code,'' Adv.

Sci. Lett., vol. 22, no. 10, pp. 2905_2909, Oct. 2016.

BIOGRAPHIES

Mrs.R.Radhika is Assistant professor in computer science and engineering department in SCSVMV university.



Buddhapuri Aneerudh is pursuing B.E(CSE) in SCSVMV university.

Bhumireddy Suneel kumar reddy is pursuing B.E(CSE) in SCSVMV university.