IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204295 |

Detection, Identification and Proactive Prevention of Phishing Websites

Rahul M [1], Sinchana K P [2], Sneha V [3], Spandana K J [4]

Prof.Raghu B R [5], Prof.Shwetha G [6]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[1]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[2]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[3]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[4]

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[5]

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[6]

ABSTRACT:Criminals seeking sensitive information construct illegal clones of actual websites and e-mail accounts. When a user clicks on a link provided by the hackers, the hackers gain access to all of the user's private information, including bank account information, personal login passwords, and images. Consumers are led to a faked website that appears to be from the authentic company when the e-mails or the links provided are opened. There are various models that are used to detect phishing Websites based on URL significance features such as Logistic Regression, Multinomial Naive Bayes, and XG Boost and many more. Various machine learning models are compared and the best one is selected based on it's accuracy .

KEYWORDS: Gradient boosting classifier, Naive bayes, XG Boost, Logic Regression

I. INTRODUCTION

Phishing is a common cyberattack technique used by cybercriminals to steal sensitive information such as login credentials, financial information, and personal data. Phishing attacks can be carried out through various means, including emails, text messages, and websites.The detection, identification, and proactive prevention of phishing websites project aims to develop an automated system that can detect and identify phishing websites in real-time and prevent users from accessing them. The system will utilize machine learning algorithms and other advanced technologies to analyze website content and detect any signs of phishing.

The project's main goal is to enhance cybersecurity by providing a reliable and efficient solution for detecting and preventing phishing attacks. By identifying phishing websites and warning users before they enter their sensitive information, the system can help reduce the risk of data breaches and financial loss.Overall, the detection, identification, and proactive prevention of phishing websites project is an important step towards strengthening cybersecurity and protecting users from the growing threat of phishing attacks.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

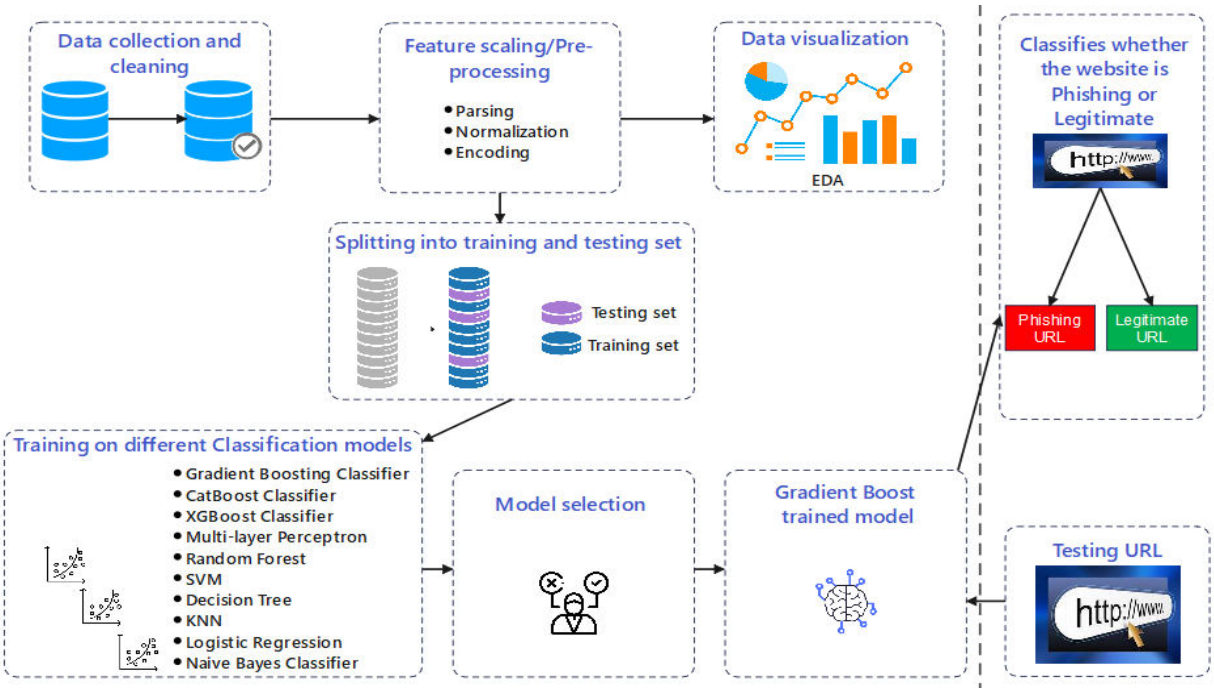
| DOI:10.15680/IJIRSET.2023.1204295 |

II. RELATED WORK

Several studies have explored the use of machine learning techniques for the identification of the phishing websites. In evaluating the e-banking phishing website, Maher Aburrous et al[1] introduced a new technique for fixing 'fuzziness' and proposed a smart, resilient, successful e- banking phishing website detection model. Their model is a mixture of flippant logic and data mining techniques to define the features of the phishing e-banking website, to analyse its techniques by categorizing phishing forms, and to define various parameters for attacking the structured e- banking phishing layer.

Xun Dong et al[2] define a new approach to the identification of phishing websites based on the study of the online activity of users ie. visited websites and submitted to those websites by data users. These user patterns cannot be openly exploited by attackers; identification based on such data can not only achieve high precision, but is also inherently resilient to changing methods of deception. Aanchal Jain et al[3] are suggesting a new approach to phishing attack detection. They have introduced a prototype web browser that can be used as an agent for phishing attacks and processes each arriving email. Using email information obtained over a period of time, they show that their method is capable of detecting more phishing attacks than current systems. PhishAri, a technique developed by Anupama Aggarwal et al[4], detects real-time phishing on Twitter. To distinguish whether a tweet posted with a URL is phishing or not, they use Twitter explicit highlights, along with URL features. Tweet content and its functionality are some of the specific Twitter parameters utilized by them. Apps such as length, hashtags, and mentions. Other Twitter parameters utilized are the attributes of the Twitter user sending the tweet, such as the number of tweets, account age, and follower-followee ratio. Sadia Afroz et al[5] suggests PhishZoo, a phishing identification technique that uses profiles of appearances of trustworthy websites to detect phishing. Phishing is a security assault involving the acquisition of confidential or otherwise private information by posing oneself as a trusted entity. Phishers also abuse the trust of users in a website's appearance by using web pages that are visually identical to an actual website.

III. METHODOLOGY



As we can see in the above figure we can see that we are going to first collect the data which will be cleaned and then we are going to preprocess them and filter them using parsing, normalization and encoding then the data will be visualized by using the EDA. The preprocessed data will be taken further and the splitting of data will be done which will be then classified into training and testing sets.This training of data will be done by using many machine learning

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

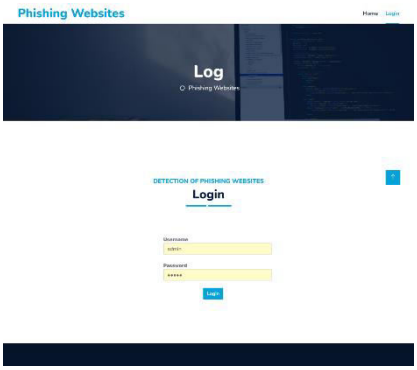
| DOI:10.15680/IJIRSET.2023.1204295 |

algorithms such as Gradient Boosting Classifier CatBoost Classifier, XGBoost Classifier, Multi-layer Perceptron,RandomForest,SVM,Decision Tree,KNN,Logistic Regression Naive Bayes Classifier but we have chosen the best algorithm among them which has the high accuracy rate that is gradient boosting classifier which gives 97 per of accuracy rate.Then with the help of this classifier the website will be classified as the phishing website or the legitimate website .

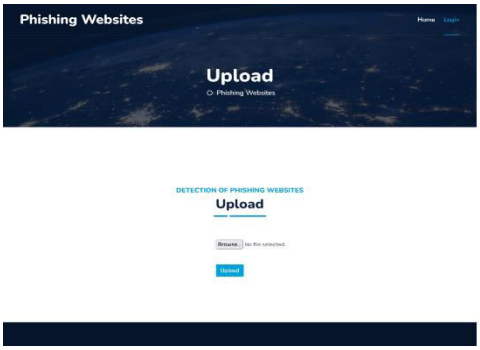
IV. EXPERIMENTAL RESULTS

Figures shows the website of Detection, identification and practive prevention of Phishing websites . (a) shows the home page of the website. (b) login page with credentials. (c) page to upload the dataset. (d) preview of the datasets.(e) here all the datasets that are uploaded will be then tested and trained.

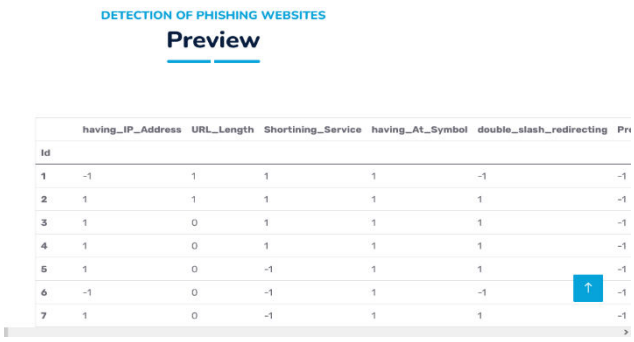
(a)



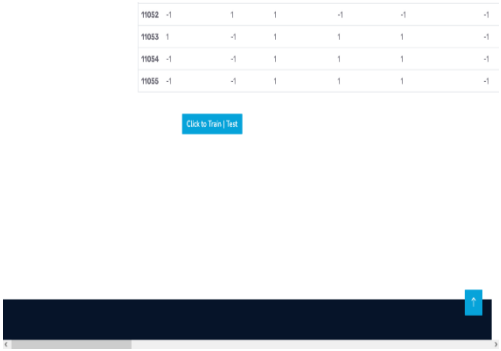
(b)



(c)



(d)



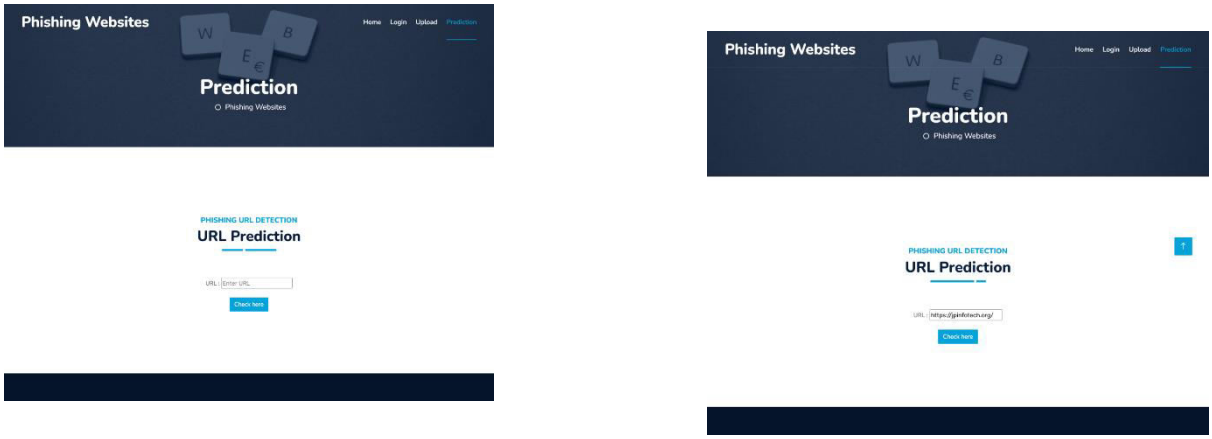
(e)

Figure 2 shows the URL predictions. (a)The page where the input URL needs to be inserted. (b) The Website's URL that needs to be classified as safe or unsafe to use is inserted.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204295 |



(a) (b)

Figure 3 shows the result obtained for the website given. (a) shows that the website is safe to use. (b) shows that the website is not safe to use. (c) It analyses the result obtained and shows the performance analysis of the URL by using recall, F1 and precision also the confusion matrix is being displayed. (d) The percentage of legitimacy and phishing in the URL is displayed in the form of pie chart.

(a) (b)

(c) (d)

V. CONCLUSION

It is remarkable that a good anti-phishing system should be able to predict phishing attacks in a reasonable amount of time. Accepting that having a good anti-phishing gadget available at a reasonable time is also necessary for expanding

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204295 |

the scope of phishing site detection. The current system merely detects phishing websites using Gradient Boosting Classifier. We achieved 97% detection accuracy using Gradient Boosting Classifier with lowest false positive rate.

REFERENCES

[1] Aburrous, Maher & Hossain, Mohammed &Dahal, Keshav &Thabah, Fadi. (2010). Intelligent phishing detection system for e- banking using fuzzy data mining Expert Systems with Applications.
[2] Dong, Xun& Clark, John & Jacob, Jerany. (2008). User behaviour based phishing websites detection.
[3] Jain, Aanchal&Richariya, Vinect. (2011). Implementing a Web Browser with Phishing Detection Techniques.
[4] A. Aggarwal, A. Rajadesingan and P. Kumaraguru, "Phish Aric Automatic realime phishing detection on twitter," 2012 eCrime Researchers Summit, Las Croabas, 2012, pp. 1-12, doi.10.1109/eCrime 2012.6489521.
[5] Afroz, Sadia &Greenstadt, Rachel. (2011). PhishZoo: Detecting Phishing Websites by Looking at Them. Proceedings - 5th IEEE International Conference on Semantic Computing, ICSC 2011. 368 - 375. 10.1109/ICSC.2011.52.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details