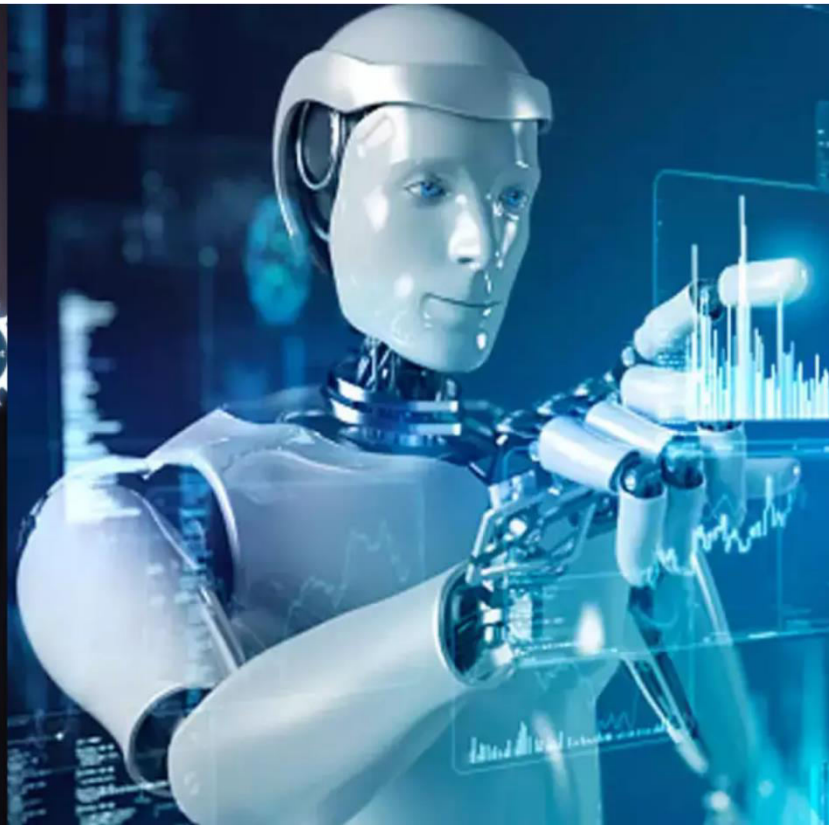
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

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

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

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

Impact Factor 8.699

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

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404136|

Prevention of Scams Using Machine Learning Models

Mrs.P.Abirami[1], Shaik Moulali[2], Mandadi Ganesh[3], M Vijaya Vardhan Reddy[4], Yallala Pedda Hussain Basha[5]

Asst. Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India[1,2]

B. Tech Students, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India[3,4,5]

ABSTRACT: In general, usage of websites is the most common things lately, it may be for ecommerce purposes or entertainment, whatever it may be. In this project, our main factor is a website, whether it is a fraudulent one or a legit one. Detection of this quality of a website is the main theme of the project. Conventionally, a website can be detected whether it is harmful or not by the browser protection service, if it is redirecting to unusual or malicious sites, such sites are marked as harmful with a symbol before the URL. Even though, the browser's firewall is enabled, it can never detect a phishing website. Because, Phishing site is not malicious site it steals data without the user even knowing it. So, to detect such sites we are training an ML model using different algorithms to determine the phishing site based on URL feature extraction. Based on different features of URL, such as like Domain length, character length etc., we will train the model with one algorithm at a time store their results and compare to find the more accurate one and display the results using the approved algorithm.

I. INTRODUCTION

- Phishing is a type of extensive fraud that happens when a malicious website act like a real one keeping in mind that the end goal to obtain touchy data, for example, passwords, account points of interest, or MasterCard numbers.
- In spite of the fact that there are a few contrary to phishing programming and methods for distinguishing potential phishing endeavours in messages and identifying phishing substance on sites, phishes think of new and half breed strategies to go around the accessible programming and systems.
- Phishing is a trickery system that uses a blend of social designing what's more, innovation to assemble delicate and individual data, for example, passwords and charge card subtle elements by taking on the appearance of a dependable individual or business in an electronic correspondence. Phishing makes utilization of spoof messages that are made to look valid and implied to be originating from honest to goodness sources like money related foundations, ecommerce destinations and so forth, to draw clients to visit fake sites through joins gave in the phishing email. The misleading sites are intended to emulate the look of a genuine organization site page.

II. LITERATURE REVIEW

MARK SOKOLOV, KEHINDE OLUFOWOBI, NIC HERNDON/2020 &Visual spoofing in content-based spam detectionWe evaluated these three scenarios with two machine learning algorithms frequently used for spam detection: decision tree and support vector machine. In This, 89% accuracy.With this approach spammers can create messages that bypass existing spam filters. Moreover, we show that this approach can be used to avoid plagiarism detection, and in other applications that use natural language processing for automatic analysis of text documents.

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

Impact Factor 8.699

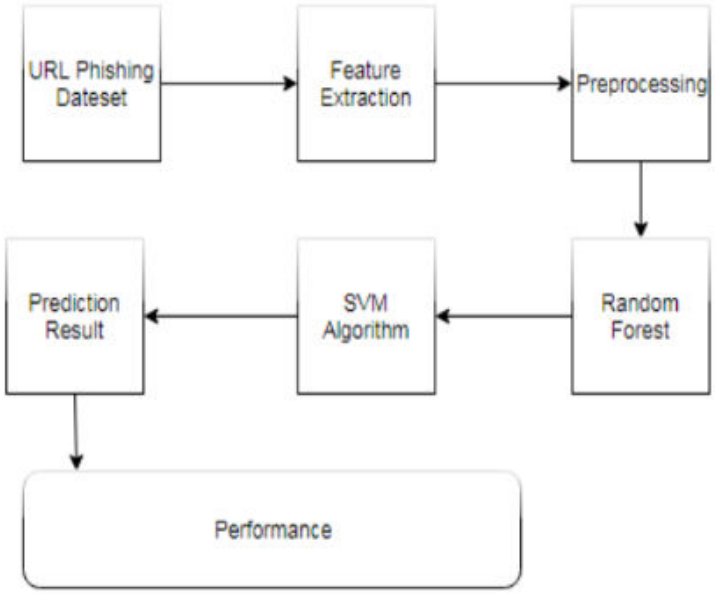
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404136|

Dr. K. Sree Ram Murthy, Dr.K.Kranthi Kumar, K.Srikar, CH.Nithya, K.Mahender Reddy/ 2020&SMS Spam Detection using RNN. We are inclined to use RNN to distinguish ham and spam sequences of variable length. Several tracked methods, such as the nearest neighbor, SVM, and neural networks, are used to counter this disadvantage.in this accuracy is 85%.The creation of inappropriate messagesfor the purpose of advertisement is harassment and the source of these messages on SMS has become the main challenge during this service. Especially SMS spam is more irritating then emailspam. We are inclined to use RNN to distinguish ham and spam sequences of variable length.

III. SYSTEM ARCHITECTURE



IV. METHODOLOGY

Existing System:

- This paper approaches a framework to extract features flexible and simple with new strategies. Data is collected from PhishTank and legitimate URLs from Google[12].
- To obtain the text properties C# programming and R programming were used.
- 133 features were obtained from the dataset and third party service providers.
- CFS subset based and Consistency subset based feature selection[12] methods used for feature selection and analyzed with WEKA tool.
- Naïve Bayes and Sequential Minimal Optimization (SMO) algorithms were compared for performance evaluation and SMO is preferred by the author for phishing detection than NB.

Proposed System:

- In this work we aim to review various machine learning methods used for this purpose, along with datasets and URL features used to train the machine learning models
- A URL based phishing attack is carried out by sending malicious links, that seems legitimate to the users, and tricking them into clicking on it. In phishing detection, an incoming URL is identified as phishing or not by analysing the different features of the URL and is classified accordingly. Different machine learning algorithms are trained on various datasets of URL features to classify a given URL as phishing or legitimate.

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

Impact Factor 8.699

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

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404136|

- we achieved 97.14% accuracy for Random Forest algorithm with the lowest false negative rate. The paper concluded that accuracy increases when more data is used for training.

Fig 1: Phishing Website Detection.

- Data Set
- Phishing Websites Features
- Detection Technique

V. IMPLEMENTATION

Data Set:

- One of the challenges faced by our research was the unavailability of reliable training datasets. In fact, this challenge faces any researcher in the field. However, although plenty of articles about predicting phishing websites using data mining techniques have been disseminated these days, no reliable training dataset has been published publically, maybe because there is no agreement in literature on the definitive features that characterize phishing websites, hence it is difficult to shape a dataset that covers all possible features.
- In this article, we shed light on the important features that have proved to be sound and effective in predicting phishing websites. In addition, we proposed some new features, experimentally assign new rules to some well-known features and update some other features.

Phishing Websites Features:

- One of the challenges faced by our research was the unavailability of reliable training datasets. In fact, this challenge faces any researcher in the field. However, although plenty of articles about predicting phishing websites using data mining techniques have been disseminated these days, no reliable training dataset has been published publically, maybe because there is no agreement in literature on the definitive features that characterize phishing websites, hence it is difficult to shape a dataset that covers all possible features.
- In this article, we shed light on the important features that have proved to be sound and effective in predicting phishing websites. In addition, we proposed some new features, experimentally assign new rules to some well-known features and update some other features.

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

Impact Factor 8.699

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

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404136|

Detection Technique:

- Detection of phishing websites has received a lot of attention recently due to their impact on users' security. Therefore, many techniques have been developed to detect phishing websites varying from communication-oriented techniques, such as authentication protocols, blacklisting, and white-listing, to content-based filtering techniques. The blacklisting and white-listing techniques have not proven though to be sufficiently efficient when used in different domains, and thus they are not commonly used. Meanwhile, the content-based phishing filters have been widely used and have proven to be of high efficiency. In light of this, researches have focused on content-based mechanism and on developing machine learning and data mining techniques based on the header and body of emails.

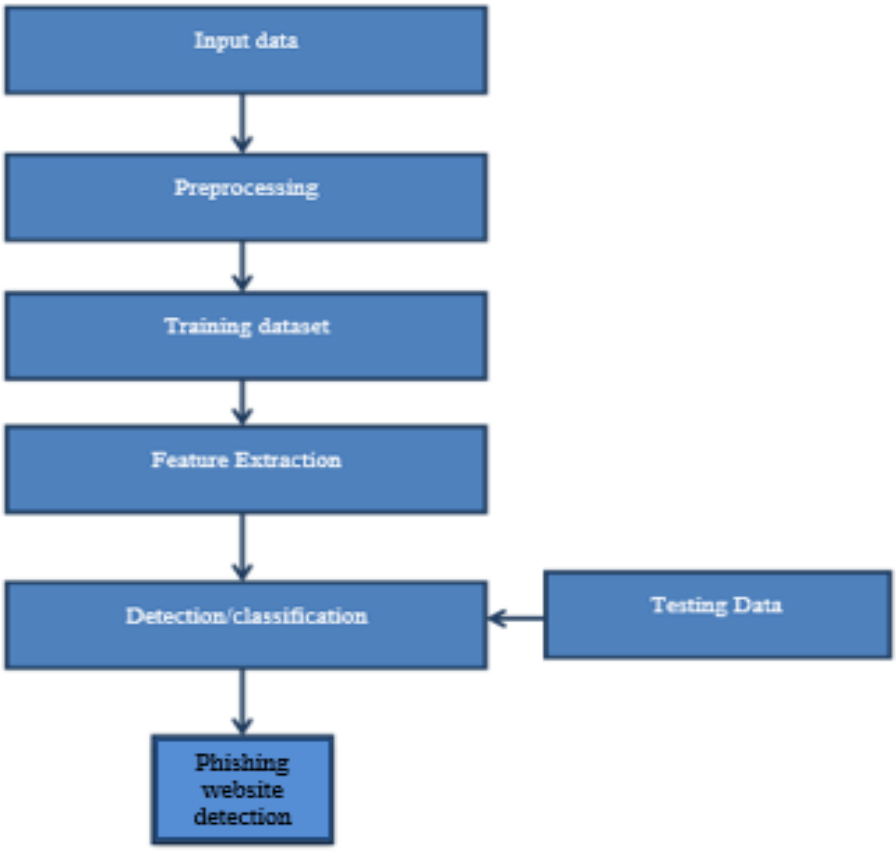


Fig 2: Data flow diagram

VI. RESULTS AND CONCLUSION

- Phishing is a cyber crime procedure utilizing both social building and specialized deception to take individual sensitive data. Besides, Phishing is considered as another extensive type of fraud. Experimentations against recent dependable phishing data sets utilizing different classification algorithm have been performed which received different learning methods. The base of the experiments is accuracy measure.

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

Impact Factor 8.699

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

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404136|

- The aim of this research work is to predict whether a given URL is phishing website or not. It turns out in the given experiment that Random forest based classifiers are the best classifier with great classification accuracy for the given dataset of phishing site. As a future work we might use this model to other Phishing dataset with larger size then now and then testing the performance of those classification algorithm's in terms of classification accuracy.

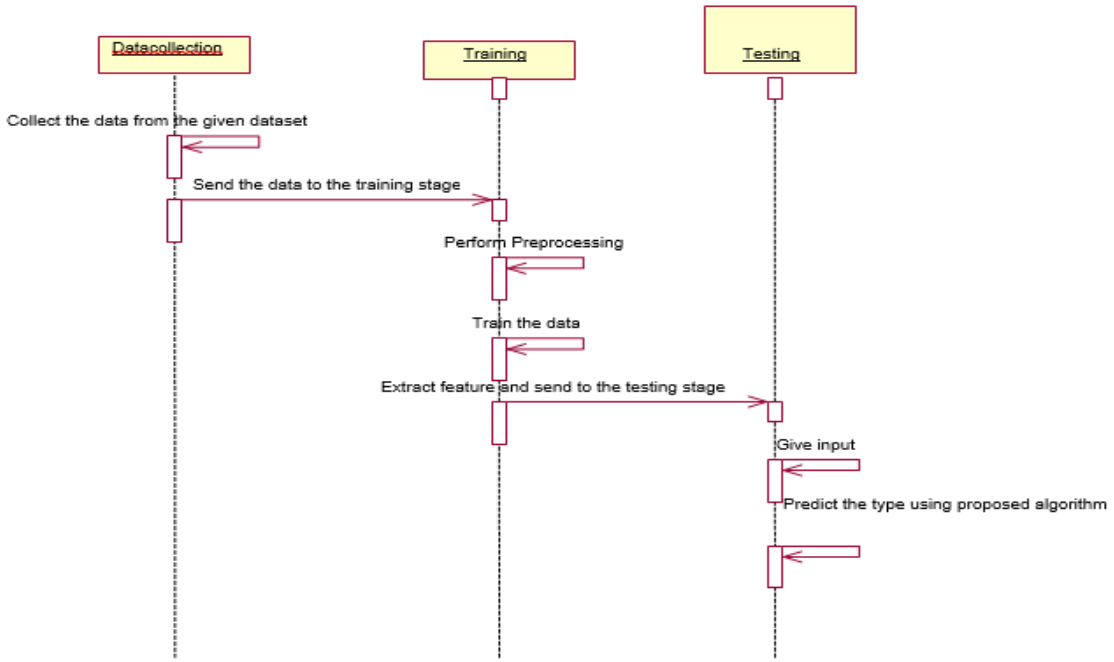


Fig 3: Sequence Diagram

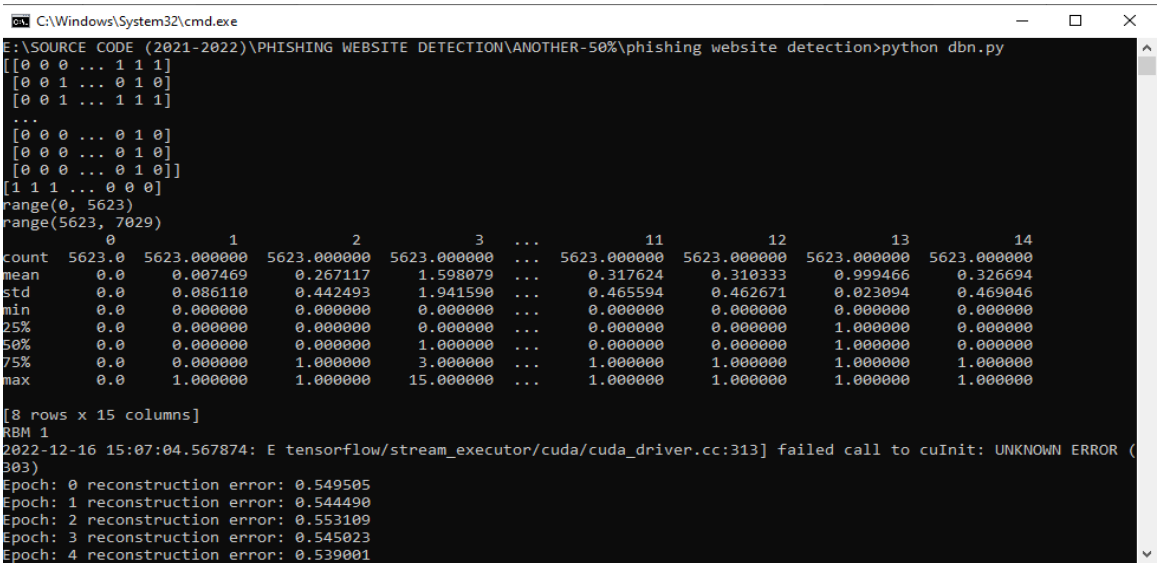


Fig 4: Performance Epoch

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

Impact Factor 8.699

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

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404136|

VII. LIMITATIONS AND FUTURE WORK

- It cannot be early predict the phishing in modern sites.
- It take more time for processing.
- Future work will aim to develop a system that can learn by itself about new types of phishing attacks by adding a more enhanced feature to the detection process.
- The scope of this approach not only helps in adding more enhanced features but also updating the existing features to improve its importance level to make detection more efficient and reduce the false positive rate to a large extent.
- Another further work should include deploying this approach into a web extension to make the detection more robust to the user.

REFERENCES

[1] Samuel Marchal, Jérôme François, Radu State, and Thomas Engel, "PhishStorm: Detecting Phishing With Streaming Analytics," IEEE Transactions on Network and Service Management, vol. 11 , issue: 4 , pp. 458-471, December 2014.
[2] Mohammed Nazim Feroz,Susan Mengel, "Phishing URL Detection Using URL Ranking," IEEE International Congress on Big Data, July 2015 .
[3] Mahdieh Zabihimayvan, Derek Doran, "Fuzzy Rough Set Feature Selection to Enhance Phishing Attack Detection," International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, June 2019.
[4] Moitrayee Chatterjee,Akbar-Siami Namin, "Detecting Phishing Websites through Deep Reinforcement Learning," IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), July 2019.
[5] Chun-Ying Huang,Shang-Pin Ma,Wei-Lin Yeh,Chia-Yi Lin,ChienTsung Liu, "Mitigate web phishing using site signatures," TENCON 2010-2010 IEEE Region 10 Conference, January 2011.
[6] Aaron Blum,Brad Wardman,Thamar Solorio,Gary Warner, "Lexical feature based phishing URL detection using online learning," 3rd ACM workshop on Artificial intelligence and security, Chicago, Illinois, USA, pp. 54-60, August 2010 .
[7] Mohammed Al-Janabi,Ed de Quincey,Peter Andras, "Using supervised machine learning algorithms to detect suspicious URLs in online social networks," IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, Sydney, Australia, pp. 1104-1111, July 2010.
[8] Erzhou Zhu,Yuyang Chen,Chengcheng Ye,Xuejun Li,Feng Liu, "OFSNN:An Effective Phishing Websites Detection Model Based on Optimal Feature Selection and Neural Network," IEEE Access(Volume:7), pp. 73271-73284, June 2019.
[9] Ankesh Anand,Kshitij Gorde,Joel Ruben Antony Moniz,Noseong Park,Tanmoy Chakraborty,Bei-Tseng Chu, "Phishing URL Detection with Oversampling based on Text Generative Adversarial Networks," IEEE International Conference on Big Data (Big Data), December 2018.
[10] Justin Ma,Lawrence K. Saul,Stefan Savage,Geoffrey M. Voelker, "Learning to detect malicious URLs," ACM Transactions on Intelligent Systems and Technology (TIST) archive Volume 2 Issue 3, April 2011.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details