IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

Antiphishing Detectors Using Machine Learning

C.Saranya[1], R.Sridevi[2], M.Subasree[3], Mrs.T.Kavitha,M.E

Assistant Professor, Department of Computer Science and Engineering, Trichy Engineering College, Konalai, Trichy, India

ABSTRACT: Malicious URLs detection techniques can be classified into Non-Machine Learning (e.g. blacklisting) and Machine learning approach (e.g. data mining techniques). Data mining helps in the analysis of large and complex datasets in order to detect common patterns or learn new things. As these web services drive increased opportunities for people to interact, they equally offer new opportunities for criminals. URLs are launch pads for any web attacks such that any malicious intention user can steal the authorized person's identity by sending the malicious URL. Malicious URLs are a keystone of unlawful Internet activities. The dangers of these sites have created a mandate for defenses that protect end-users from exploring them. The proposed approach is that which classifies URLs automatically by using a Machine learning algorithm called logistic regression that uses binary classification. Training dataset will generate a model which is used to predict the testing dataset of 15,56,896 URLs. Data mining techniques can generate classification models which is used to manage data, modelling of data that helps to make prediction about whether it is malicious or legitimate. In this project we achieve accuracy of 98.7% detecting malicious URLs.

KEYWORDS: Phishing Detection, Machine learning algorithm-Logistic regression, Multinomial NB.

I. INTRODUCTION

Phishing has become one among the most deadly attacks. There are various approaches to thwarting phishing attacks from Associate in society. To avoid this type of attacks ,we are introducing Anti Phishing Technology in order to ensure the higher Security.Phishing is an online identity theft in which an attacker uses fraudulent e-mails and bogus website in order to trick gullible customers into disclosing confidential information such as bank account information, website login information, and so forth. (Topkara et al., 2005). Phishing is an indicative type of illegal fraudulent attempt in online electronic communication. Phishing is a form of internet scam in which an attacker makes use of an email or website to illegally obtain private (Martin et al., 2011). It is a semantic attack which aims at harming the user rather than the computer. In general, phishing is a relatively new internet crime. The ease of cloning a legitimate bank website to convince unsuspecting users has made phishing difficult to curtail. Mostly, an email with a redirecting website link is sent to the user to update confidential information such as credit card, website login information, and bank account information that belongs to the licit. As explained by Aburrous et al. (2008), the complexity of understanding and analysing phishing website is as a result of its involvement with technical problems and social. The main effect of phishing website is in the abuse of information through the compromise of user data that may harm victims in form of 2 A Machine Learning Approach to Phishing Detection and Defence financial losses or valuables. Phishing in comparison to other forms of internet threat such as hacking and virus is a fast-growing internet crime.

PROBLEM STATEMENT

Phishing is an online identity theft in which an attacker uses fraudulent e-mails and website in order to trick customers into disclosing confidential information such as bank account information , website login information, and so forth. Phishing is a form of internet scam in which an attacker makes use of an email or website to illegally obtain private.
Anti-phishing tools to detect malicious URL in an organization to protect its users. In the event of malicious code being implanted on the website, hackers may steal user information and install malware, which poses a serious risk to cybersecurity and user privacy

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

II. OBJECTIVE

The detection and mitigation of phishing attacks is a grand challenge due in the real world . There have been numerous studies on detecting and mitigating Phishing attacks. Anti Phishing Technology is an effective and efficient solution to detect and mitigate phishing attacks with its accuracy of 98.7%.

1.Project description

1.1 Different types of phishing

In the broad usage of internet as a major form of communication, phishing can be implemented in different ways such as follows (Alnajim and Munro, 2009):

Email-to-email: when someone receives an email requesting sensitive information to be sent to the sender.
Email-to-website: when someone receives an email embedded with phishing web address.
Website-to-website: when someone clicks on phishing website through a search engine or an online advert.
Browser-to-website: when someone misspelled a legitimate web address on a browser and then referred to a phishing website that has a semantic similarity to the legitimate web address.

1.2 URL Detection Techniques

Phishing in modern society is highly urgent, challenging, and overly critical. There have been several recent studies against phishing based on the characteristics of a domain, such as website URLs, website content, incorporating both the website URLs and content, the source code of the website and the screenshot of the website. However, there is a lack of useful anti- phishing tools to detect malicious URL in an organization to protect its users. In the event of malicious code being implanted on the website, hackers may steal user information and install malware, which poses a serious risk to cybersecurity and user privacy. Malicious URLs on the Internet can be easily identified by analyzing it through Machine Learning (ML) technique. In comparison to most previous approaches, researchers focus on identifying malicious URLs from the massive set of URLs. Therefore, the study proposes Recurrent Neural Network (RNN) based URL detection approach. The objectives of the study are as follows:

- To develop a novel approach to detect malicious URL and alert users.
- To apply ML techniques in the proposed approach in order to analyze the real time URLs and produce effective results
- To implement the concept of RNN, which is a familiar ML technique that has the capability to handle huge amount of data.

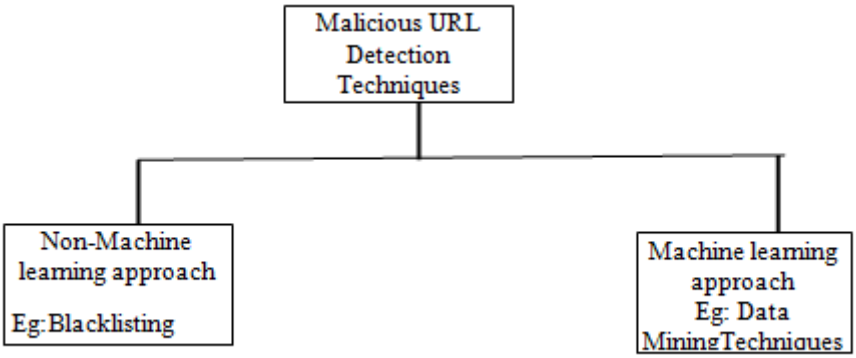


Fig 1 Detection techniques

1.3 Machine learning models :The machine learning model is nothing but a piece of code; an engineer or data scientist makes it smart through training with data. So, if you give garbage to the model, you will get garbage in return, i.e. the trained model will provide false or wrong predictions.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

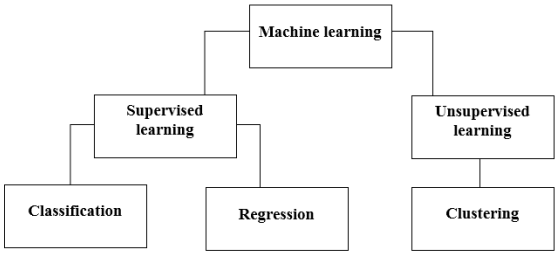


Fig 2 Machine learning models

II. LITERATURE SURVEY

1.Abutair et al. (2019) introduced a case-based reasoning approach for phishing detection and named it CBR-PDS. They have used feature extraction and blacklisting of URLs and a genetic algorithm weighing technique to predict the phishing URLs.
2.Sonowal and Kuppusamy (2020) have come up with their developed model named PhiDMA. It is a multi-filter model comprised of five layers—A whitelist layer, features layer, lexical signature, string matching, and a comparison layer for accessibility score. They also introduced their algorithm to predict phishing website URLs.
3.Aassal et al. (2020) have proposed a benchmarking and evaluation of phishing detection research studies, and the model was named PhishBench. In the proposed approach, they used term frequency-inverse document frequency (TF-IDF)
Orunsolu et al. (2020) have proposed a feature selection-based approach using Support Vector Machines and 4.Naïve Bayes machine learning classifiers using the weka tool to predict phishing URLs.
5.Hong et al. (2020) introduced an approach to perform phishing URL detection using lexical features and blacklist domains using Adaboost, Random Forest, and SVM. They have also used 3 string-based classifiers; and deep learning classifiers: 1DConv, LSTM, and 1DConv + LSTM

III. METHODOLOGY

1. Collection of both malicious URLs and legitimate URLs
2. Extract the features of URLs to detect the malicious and legitimate URLs.
3. Tokenize the words, and stem the data using Snowball Stemmer .
4. Creation of training dataset and testing dataset on the basis of features extracted.
5. Training using Logistic Regression and MultinomialNB for finding accuracy.
6. In testing, we found Logistic Regression with 98% of accuracy and trained the model using the same, by pipeline the data.

Let be the set of URLs where m is the maximum limit for the number (n) of URLs. Let M, L ∈ xn be the malicious and legitimate, accordingly. Suppose M and L contains the properties Pm and Pl, respectively.
Malicious = Output(Input(Pm)) Legitimate Output(Input(PI))
Cell state (CS)-It indicates the cell space that accommodate both long term and short-term memories.
Hidden state (HS)-This is the output status information that user use to determine URL with respect to the current data, hidden condition and current cell input. The secret state is used to recover both short-term and long-term memory, in order to make a prediction.
Input gate (IT)-The total number of information flows to the cell state. er of I
Forget gate (FT)-The total number of data flows from the current input and past cell state into the present cell state.

IV. EXISTING SYSTEM

Non-content-based approaches: It include URL and host information-based classification of phishing sites, blacklisting, and whitelisting methods. In URL-based schemes, URLs are classified on the basis of both lexical and host features. Lexical features describe lexical patterns of malicious URLs. These include features such as length of the URL, the number of dots, special characters it contains. Host features of the URL include properties of IP address,the owner of the site, DNS properties such as TTL Using thesefeatures, a matrix is built and run through multiple classification algorithms. In real-time processing trials, this approach has success rates between 95% and 99%.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

Content-based approaches: The classifier is frequently retrained with new phishing sites to learn new trends in phishing. This classifier has high accuracy but is presently implemented offline as it takes 76 seconds on average to detect phishing. Some re-searchers studied fingerprinting and fuzzy logic-based approaches that use a series of hashes of websites to identify phishing sites.A poorly structured NN model with random forest algorithm may cause the model to under fit the training dataset . On the other hand, It will take time to load all the dataset. Process is not accuracy. It will analyse slowly and exaggeration in restructuring the system to suit every single item in the training dataset may cause the system to be over fitted. This Model has been worked on emails .

Disadvantages:
1.It will take time to load all the dataset.
2.Process is not accuracy.
3.It will analyse slowly.

V. PROPOSED SYSTEM

Proposed system Here we proposed a new method of anti-phishing technology. The Anti-Phishing Technology using Machine Learning Approach is a mechanism that is proposed in order to ensure high security. In this mechanism we deal with the URLs (Uniform Resource Locaters) and the URI (Uniform Resource Identifies) check with machine learning technique and predict whether it is phishing website or not. Here we create a web app for browsing imputed URLs. Each time we browse a site the corresponding URL (Uniform Resource Locater) of site will be checked with machine learning technique. Phishing site detection is truly an unpredictable and element issue including numerous components and criteria that are not stable. On account of the last and in addition ambiguities in arranging sites because of the intelligent procedures programmers are utilizing, some keen proactive strategies can be helpful and powerful tools can be utilized, for example, fuzzy, neural system and data mining methods can be a successful mechanism in distinguishing phishing sites. We applied logistic regression algorithms to model our train out model and at the end logistic regression which gave a more accurate prediction was used in our system. 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. 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

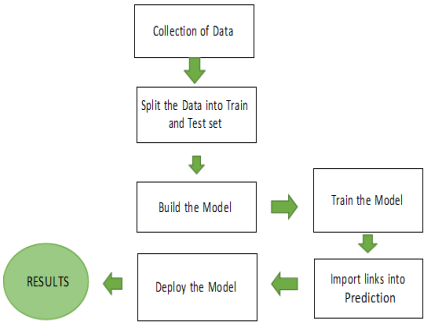


Fig 3 Architectural model

Advantages:

- It has Better Accurancy .
- We used supervised learning using multinomial navie bayes algorithm, logical regression by vectorization the data set.
- All of URLs in the dataset are labelled.
- The Data Set is taken from Phishtank .

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

VI. IMPLEMENTATION

6.1 Preprocessing: Standardization of datasets is a common requirement for many machine learning estimators implemented in scikit-learn; they might behave badly if the individual features do not more or less look like standard normally distributed data: Gaussian with zero mean and unit variance.

6.2 Regexptokenizer: A regex based tokenizer that extracts tokens either by using the provided regex pattern (in Java dialect) to split the text (default) or repeatedly matching the regex (if gaps is false). Optional parameters also allow filtering tokens using a minimal length. For example, the following tokenizer forms tokens out of alphabetic sequences, money expressions, and any other non-whitespace sequences:>>> from nltk.

6.3 Snowball Stemmer: Snowball is a small string processing language for creating stemming algorithms for use in Information Retrieval, plus a collection of stemming algorithms implemented using it. Stemming maps different forms of the same word to a common "stem" - for example, the English stemmer maps connection, connections, connective, connected, and connecting to connect. So a searching for connected would also find documents which only have the other forms. This stem form is often a word itself, but this is not always the case as this is not a requirement for text search systems, which are the intended field of use. We also aim to conflate words with the same meaning, rather than all words with a common linguistic root (so awe and awful don't have the same stem), and over-stemming is more problematic than under-stemming so we tend not to stem in cases that are hard to resolve. If you want to always reduce words to a root form and/or get a root form which is itself a word then Snowball's stemming algorithms.

6.4 Visualization: Word Cloud represents your text data as a cluster of words, where the importance of the words is shown with font size and color. Many time's you might have seen a cloud filled with lots of words in different sizes, which represent the frequency or the importance of each word. This is called a Tag Cloud or word cloud. It's important to remember that while word clouds are useful for visualizing common words in a text or data set, they're usually only useful as a high- level overview of themes. They're similar to bar blots but are often more visually appealing (albeit at times harder to interpret). However, it's important to keep in mind that word clouds don't provide any context or deeper understanding of the words and phrases being used. Therefore, they should be used in conjunction with other methods for analyzing and interpreting text data.

6.5 Model selection: Machine learning is a subsection of artificial intelligence (AI) that offers systems the skill to mechanically learn and improve from experience without being explicitly programmed. Machine learning concentrates on the development of computer programs that can access data and use it learn for themselves. The procedure of learning begins with data, such as examples, direct understanding, or instruction, in order to look for outlines in data and make better conclusions in the feature based on the examples that provide. The main aim is to permit the computers to learn automatically without human interference or assistance and regulate actions consequents

6.6 Logistic regression: Logistic Regression is a classification model that is used when the dependent variable (output) is in the binary format such as 0 (False) or 1 (True). This makes logistic regression a good algorithm fit for the purpose of our work in predicting if a URL is a phishing URL (1) or not (0) as in the case of this paper. Logistic Regression is an extension of the Linear Regression model. Let us understand this with a simple example. If we want to classify if an email is a spam or not, if we apply a Linear Regression model, we would get only continuous values between 0 and 1 such as 0.4, 0.7 etc. On the other hand, the Logistic Regression extends this linear regression model by setting a threshold at 0.5, hence the data point will be classified as spam if the output value is greater than 0.5 and not spam if the output value is lesser than 0.5. In this way, we can use Logistic Regression to classification problems and get accurate predictions. The logistic function, also called as sigmoid function was initially used by statisticians to describe properties of population growth in ecology. The sigmoid function is a mathematical function used to map the predicted values to probabilities. Logistic Regression has an S- shaped curve and can take values between 0 and 1 but never exactly at those limits. The logistic function is defined as: **Logistic(n) = 1/(1+exp(-n)).**
Logistic Regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X

6.6.1 Confusion Matrix with using Logistic Regression: The confusion matrix is in the form of a square matrix where the column represents the actual values and the row depicts the predicted value of the model and vica versa.The

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

confusion matrix is a tool for predictive analysis In machine learning. In order to check the performance of a classification based machine learning model, the confusion matix is deployed.Also we can say Confusion matrix is a summerized table of the number of correct and incorrect predictions yielded by a classifier (or a classification model) for binary classification tasks. • A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. By visualzing the confusion matrix, an individual could determine the accuracy of the model by observing the diagonal values for measuring the number of accurate classification.

After that learn them from data using Logistic Regression and to accuracy goes to 98.7% and it is the good performance than before it shown below the Figure

Fig 4 **Logistic regression performance**

6.7MultinomialNB: Naïve Bayes uses assumption that the Xi are conditionally independent, given Y
P(X1, X2Y) = P(X1 X2,Y)P(X2|Y) *Chain rule
= P(XY)P(X2Y) *Conditional Independence in general: P(X1...XnY) = [] P(X,Y) .
Combining probability distribution of *PP* with fraction of documents belonging to each class, given in

Pr(c) a π II Pr(wic)/w
In order to avoid underflow, we have used the sum of logs as in
Pr(c) a log(x, I, Pr(wic)/w) Pr(c)
= log πe+Ilfw log(Pr(wc))
One issue is that, if a word appears again, the probability of it appearing again goes up. In order to smooth this, we take the log of the frequency
Pr(c) = logπ+I, log(1+f) log(Pr(w|c))
Also, in order to take stop words into account, we have added here an Inverse Document Frequency (IDF), tt is the term frequency in the document dd, the word tt number of times term tt appears in a document dd in (Eq. 8) and for further process put in
Pr(c)=logn,+If log(t, Pr(wic))
Even though the stop words have already been set to 0 for this specific use case, the IDF implementation is being added to generalize the function. As we can see, IDF has little effect as we removed the stop words. However, for the smoothing it makes the model more accurate. Hence, our optimal model is mathematically expressed in Pr(c) = log π, +ΠI, log(1+f) log(Pr(wc)) .

6.7.1Confusion Matrix with using MultinomialNB: The confusion matrix is in the form of a square matrix whre the column represents the actual values and the row depicts the predicted value of the model and vica versa. After that learn them from data using MultinomialNB and to accuracy goes to 98.2% and it is the good performance than before.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

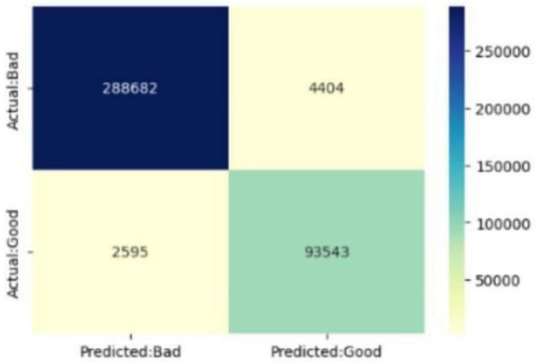


Fig 5 Multinomial Performance

6.8 Phishing dataset: Phishtank is a familiar phishing website benchmark dataset which is available at https://phishtank.org/. It is a group framework that tracks websites for phishing sites. Various users and third parties send alleged phishing sites that are ultimately selected as legitimate site by a number of users. Thus, Phishtank offers a phishing website dataset in real-time. Researchers to establish data collection for testing and detection of Phishing websites use Phishtank's website. Phishtank dataset is available in the Comma Separated Value (CSV) format, with descriptions of a specific phrase used in every line of the file. We use dataset from various sources, and combined all together in list of CS

6.9 Vectorization: Machine learning algorithms operate in a space of numerical attributes, that is, they expect that a two dimensional array will be presented at the input, the rows of which are concrete instances, and the columns are attributes or features. Thus, in order to perform machine learning on the text, it is necessary to convert the source documents into vector representations, to which numerical machine learning will subsequently be applied. This process is called vectorization and it is the first step towards analyzing natural language data. Converting documents to their numerical form makes it possible to analyze them and create instances with which the machine learning algorithm we choose will work. Documents (or sentences) can have different sizes, but the vectors that we define for them will always be the same length. Each property in a vector representation is a feature. In our case, these will be the words that are included in the sentence. Together, all these features will describe a multidimensional feature space to which machine learning methods can be applied. Thus, we must move from individual sentences and words to points in a multidimensional semantic space. These points can be located far or close to each other, distributed evenly or vice versa randomly. Based on this, we can conclude that sentences that are close in meaning will be located nearby, and different, on the contrary, far.

6.10 Frequency vectorizer: One way to vectorize the source text is to calculate the frequency of occurrence of each word in each sentence and associate this value with the entire set of words of the original data set. You can start by creating a dictionary of all words in all sentences of your dataset. A dictionary in this case is a list of words that occur in texts where each word has its own index. This allows us to create a vector for any sentence - just take the sentence that we want to vectorize and count the occurrence of each word. The length of the resulting vector will be equal to the size of our dictionary and contain the value of the number of occurrences of the word from the dictionary in each specific sentence. Then we can use the method CountVectorizer from the scikit-learn library to vectorize our sentences. The result of the CountVectorizer() vectorizer. As the output we get a dictionary of all unique words that are available in all sentences that we passed as the input to our vectorizer. We can also take all our sentences and transform them using CountVectorizer() to get vectors for each sentence that will display the number of occurrences of each word from the dictionary in a specific sentence.

6.11 Count Vectorizer: Count vectorizer makes it easy for text data to be used directly in machine learning and deep learning models such as text classification. Count Vectorizer is a great tool provided by the scikit-learn library in Python. It is used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text. This is helpful when we have multiple such texts, and we wish to convert each word in each text into vectors (for using in further text analysis). Count Vectorizer creates a matrix in which each unique word is represented by a column of the matrix, and each text sample from the document is a row in the matrix. The value of each cell is

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

nothing but the count of the word in that particular text sample. Inside Count Vectorizer, these words are not stored as strings.Rather, they are given a particular index value. In this case, ‘at’ would have index 0, ‘each’ would have index 1, ‘four’ would have index 2 . This is helpful when we have multiple such texts, and we wish to convert each word in each text into vectors (for using in further text analysis). Rather, they are given a particular index value. In this case, ‘at’ would have index 0, ‘each’ would have index 1, ‘four’ would have index 2 and so on

VII. SYSTEM DESIGN

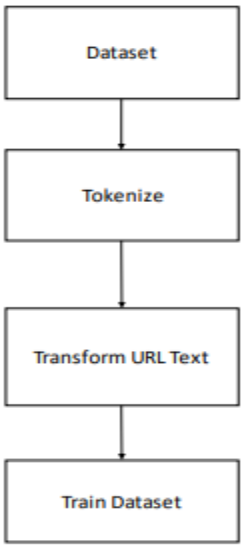
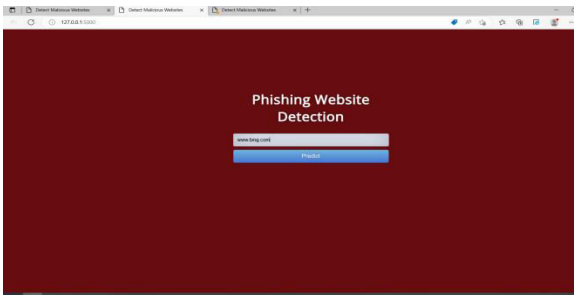


Fig 6 Data flow diagram-1

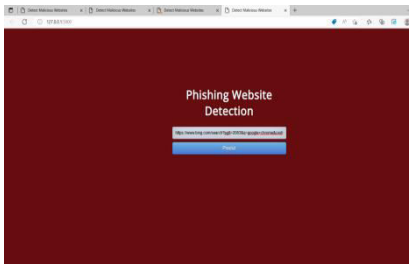
Fig 7 Data flow diagram-2

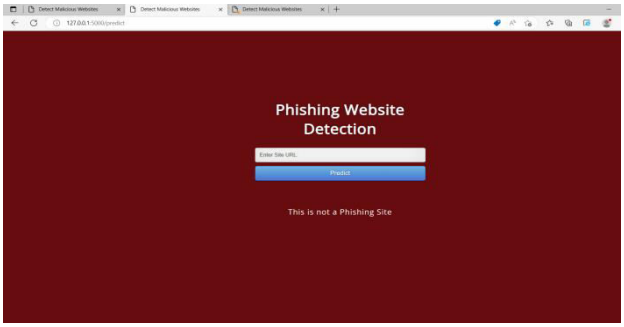
VIII. RESULTS

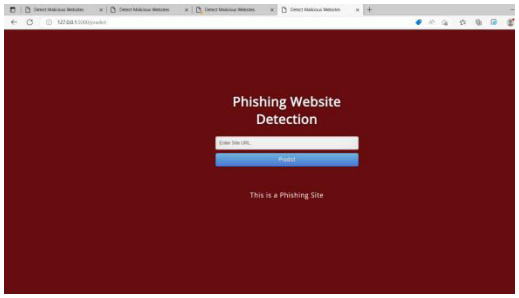
Non Phishing site



Phishing site







| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105079 |

IX. CONCLUSION

Due to the growing use of Internet in our daily life, cyber attackers aim their victim over this platform. One of the mostly encountered attack is named as "phishing" which creates a spoofed web page to obtain the users sensitive information such as user-ID and password in financial websites by using social networking facilities. The malicious web page is created as if a legitimate web page, especially copying the original web page one to one. Therefore, detection of these pages is a very trivial problem to overcome due to its semantic structure which takes the advantage of the humans' vulnerabilities. Software tools can only be used as a support mechanism for detection and prevention this type attacks, and these tools especially use whitelist/blacklist approach to overcome this type of attacks. However, they are static algorithms and cannot identify the new type of attacks in the system. Therefore, as an efficient solution, we propose the use of logistic regression machine learning system for classifying the incoming URLs. The experimental results show that this approach result satisfactory accuracy rate of about 98.7% of accuracy.The latency of the execution time of the algorithm is also an important metric for selection of the detection algorithms. As seen from the results use of Alexa Ranking results a great increase in the execution time, although it has a great importance for detection of phishing.

X. FUTURE ENHANCEMENT

As the Future works, to decrease the execution time and increase the efficiency of the system, the power of the Graphics Programming Units can be used. Additionally, other approaches of Deep Learning, such as recurrent neural networks and convolutional neural networks can be tested for increasing the performance of the system.

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. 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_4
3. Purbay M., Kumar D, “Split Behaviour of Supervised Machine Learning Algorithms for Phishing URL Detection”, Lecture Notes in Electrical Engineering, vol. 683, 2021, Doi: 10.1007/978-981-15-6840-4_40 [CrossRef] [Google Scholar]
4. Y.Sönmez, T. Tuncer, H. Gökal, and E. Avci, “Phishing web sites features classification based on extreme learning machine,” 6th Int. Symp. Digit. Forensic Secure. ISDFS 2018 - Proceeding, vol.
5. 2018–Janua, pp. 1–5, 2018.
6. R.Kiruthiga, D. Akila,” Phishing Websites Detection Using Machine Learning” International Journal of Recent Technology and Engineering (IJRTE) ISSN:2277- 3878, Volume-8, Issue -2S11, September 2019
7. 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).
8. PMCID: PMC8504731 PMID: 34634081
9. 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,
10. arXiv:1802.03162, July 2017.
11. 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.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details