JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Malicious URL's Detection Using Machine Learning

[1] P. Sai Sireesha, [2] K. Padmini,[3] B. Balu, [4] B. Surya Teja, [5] U D Prasan

[1]Student, [2] Student, [3] Student,[4] Student,[5] Guide
[1] Department of Computer Science & Engineering
[1] Aditya Institute of Technology and Management, Tekkali-532201, India

Abstract: The proliferation of cybersecurity threats, particularly in the form of malicious websites, has become a pressing concern with the widespread use Software Devices for Various purposes. Malicious websites host content ranging from spam and malware to inappropriate ads and spoofing, posing a significant risk to users who may fall victim to scams, resulting in financial loss, private information disclosure, malware installations, and other detrimental consequences. The financial toll of such incidents amounts to billions of rupees each year. Traditional detection methods, primarily reliant on blacklists, prove insufficient in addressing the evolving nature of cyber threats, particularly in identifying newly generated malicious URLs. This project seeks to address this gap by leveraging various machine learning algorithms to proactively detect and classify malicious URLs. Treating the problem as a multi-class classification challenge, raw URLs are categorized into different types, including benign or safe URLs, phishing URLs, malware URLs, and defacement URLs using various machine learning algorithms. The primary objective is to reduce cyberattacks by preventing users from clicking on malicious URLs, thus averting potential financial and informational losses.

KEYWORDS: Malicious, URL's, Random Forest, Machine Learning, Classification, Phishing.

1)INTRODUCTION

1.1 URL S and Malicious URL's:

A URL is like an address for things on the internet. It has two main parts: one that says how to access it (like using "http" or "https"), and another that shows where it's located (like the website's name or IP address). Because URLs follow a specific structure, bad actors can try to change parts of it to create harmful URLs. These malicious URLs might trick people into going to unsafe websites, running harmful code on their computers, or downloading harmful software. So, it's important to be cautious and check URLs to stay safe online. It consisting several components such as Scheme, Subdomain, Domain, Top level Domain, Port number, Path, Query String Separator and Parameter, Fragment.

1.2 Technology's:

At first blacklisting system is like a protection for our online safety. It works by keeping a list of bad websites called a blacklist and making sure we don't visit them. This system is always learning and updating its list to include new threats. It collects information from different places, like cybersecurity experts and special tools that search the internet for bad stuff. By looking for patterns that show a website is up to no good, the system creates a list of websites to avoid. This list is then used in security tools, like antivirus programs and web browsers. When we try to open a website, the blacklisting system quickly checks if it's on the bad list but it has few limitations.

Finding bad websites is really important nowadays, and scientists are using smart computer techniques to do it. Social media sites like Twitter and Facebook, where people share web links a lot, are especially at risk. Studies show that about 25% of messages on these platforms have web links, making it hard to handle all the possibly dangerous stuff. Because of the COVID-19 pandemic, more businesses are online, but this also means more cyber threats, like harmful websites trying to trick people. So, it's crucial to figure out which websites are bad to keep everyone safe online.

Detecting bad websites is important for different industries. It helps protect people's money, makes the internet safer, and lets web browsers warn users about fake websites. Special filters also help stop fake emails. The new system, which tells if a website is harmful or not, helps measure how well it works by comparing results from different models. Malicious URL detection is instrumental in distinguishing between benign and harmful online activities, including phishing, malicious endeavors, and defacement. While phishing involves deceptive websites tricking users into revealing sensitive information, benign phishing might refer to unintentional or harmless attempts that resemble phishing but lack malicious intent.

On the other hand, malicious URLs play a critical role in identifying and blocking harmful websites that aim to compromise user data or inject malicious code into devices. Defacement, the alteration of a website's appearance, can have both benign and malicious motives. Malicious URL detection helps pinpoint and prevent such alterations, safeguarding websites from digital vandalism and maintaining the trust and security of online platforms. The constant evolution of malicious tactics makes robust URL detection an essential component in preserving the integrity of online spaces and protecting users from potential threats.

Abbreviations and Acronyms (Heading2)

URL: Uniform Resource Locator
HTTP: Hypertext Transfer Protocol
ML: Machine Learning
RF: Random Forest
TLD: Top-Level Domain
XGB-Extreme gradient boost

2) 2.1 Literature survey :

Over the past few years, multiple methodologies have been proposed and employed for detecting malicious URLs.

1. Research on Malicious URLs detection was done by Shantanu Maheshwari [1] in the year 2021. This concludes that the using of Machine learning algorithm’s for detecting URLs have decreased the risk of attackers. The data model covers various stages like data collection, data processing and data visualization steps. This research using algorithm’s KNN , Gaussian Naïve bayes, Stochastic Gradient Descent The output is compared accuracy of all the predictions of the above algorithm’s. This produces a Higher accuracy of KNN as the 94%.
2. In the year 2021, Gold Wejinya and Sajal Bhatia [2] Has done research on Machine Learning for Malicious URL’s Detection which was published in the book ICT Systems and sustainability. It uses Comparison of algorithm’s which include SVM (support vector machine), Navie Bayes, Logistic Regression Methods. In this Particular Research the Navie bayes produced an output accuracy as more predicted than another algorithm’s. In this particular research it contains various pre-processing steps and evaluation methods which include f1 score, accuracy, precision, recall. The results of the compared outputs are treated .
3. A lexical features-based selection was proposed in the year 2022 using Deep learning-based techniques like multi-layer perception and Decision Tree by Warmn Ahmed and Noor Ghazi M. Jameel [3]. During his study they took the taxonomy of Detecting of Phishing URLS on Arabic and English Websites using machine learning. Based on the Language the classified into English and Arabic. The features take lexical and content features and content features. This paper focus on reviewing studies About detecting of malicious URLs using Algorithm’s considering Arabic and non-Arabic contents.
4. Detecting Malicious URL’s is research conducted by K. Rohit Verma. Manmohana K, Immadisetti Naga Venkata Durga Naveen [4] in year 2020, which uses the method feature representation of the URL. Here they followed the Lexical Analysis Features in addition to 3rd party features. Classification Based on TF-IDE word association. They used a feature is 16 safe browsing which is a binary valued and it ‘1’ indicate the begin and ‘0’ indicates the malicious. The 17 Alexa party services are used in rank host feature.
5. In the year 2020, Vara Vandavall, Farhat Lamia Barsha, Mohammad Masum, Hossain Shahriar, Hisham Haddad [5], has done research on Malicious URL Detection using Supervised machine Learning Techniques which was published in international Conference on Security of Information and Networks .In this particular research we observe they used to gathering URL using dataset and assemble URLs and compares Logistic Regression, Naïve Bayes, and Neural Networks techniques to identify malicious websites. In this process they concluded that Naïve Bayes produces more accuracy than Logistic Regression and Neural Networks.
6. Research on malicious URL Detection Based on Associative Classification was done by Sandra Kumi, ChaeHo Lim and Sang-Gon Lee [6] in year 2021.They using CBA (Classification Based on Association) algorithm to classify the malicious URLs published in entropy MDPI. The CBA algorithm uses a training dataset of URLs as historical data to discover association rules to build an accurate classifier and also, they compare with other classification models like Precision, Recall, ROC Area, False Positive Rate they got 95% accuracy.
7. In the year 2022, Amogh, Rajendra, Deekshith, S Parikshith, Asst. Prof. Suma [7], has done research on Malicious URL Detection using Machine learning and Deep Learning Which are published in International Research Journal of Modernization in Engineering Technology and Science. Here they used CNN (Convolutional Neuron Networks) &RNN (Recurrent Neuron Network) to classify the malicious URLs. The user will interact with system through GUL (Graphical User Interface). The new user should login, after that he can upload a picture or text or link than the CNN will extract features or process the images by preprocessing and produce the description about the link to the user.
8. Detecting Malicious URL is research on 2019 by BBC. Andrea Turiaková[8] published in Masaryk University Faculty of Informatics .In this research they used SVM(Support Vector Machine) and TensorFlow and found to get similar results .SVM is faster on small amount of data but time consumption is more whare as TensorFlow time consumption is consistent and more time efficient solution.
9. In research of Malicious URL Detection based on Machine Learning was conducted by Cho Do Xuan, Hoa Dinh Nguyen, Tisenko Victor Nikolaevich [9] published in International Journal of Advanced Computer Science and Applications in 2020.By using feature extraction, feature selection classify the malicious URLs machine learning techniques based on our proposed URL behaviors and attributes.
10. A research article on a Malicious URL Detection Model Based on Convolutional Neural Network done by Zhiqiang Wang, Xiaorui Ren, Shuhao Li, Bingyan Wang, Jianyi Zhang and Tao Yang [10] in 2021 was published in Hindawi Security and Communication Networks. In this research they design a malicious URL detection model based on a dynamic convolution neural network (DCNN) to solve these problems and they produce higher accuracy of CNN as 98%.
11. In the year 2022, a research on Malicious Website Detection using Machine Learning done by Mahek Khera, Tanishka Prasad, Liudmila Swati Xess , Rupender Singh, Manpreet Kaur Aiden was published in International Journal Of

Engineering Research & Technology (IJERT).they gathered a information on the URL to provide additional information like port status, directories and subdomains associated with the website. Moreover, they propose a machine learning approach to detect malicious websites using the machine learning model with best accuracy.

12. In the year 2023, N. Venkatesh, V. Tejaswini, G. Soumya,T. Shiva Priya onMalicious URL Detection Using Machine Learning was Published in Turkish Journal of Computer and Mathematics Education (TURCOMAT) . Phishing detection is a challenging problem, and many different solutions are proposed in the market as a blacklist, rule-based detection, anomaly-based detection by using this method they classify malicious URLs.

3) METHODOLOGIES:

3.1 PROPOSED SYSTEM:

In the initial stages of malicious URL detection, conventional methods relied on blacklisting and heuristic techniques. However, these approaches presented limitations, especially with the rapid advancement of technology leading to a surge in the daily generation of vast numbers of links. As a consequence, these methods struggled to keep pace with the evolving landscape of malicious URLs.

Recognizing the need for more adaptive and efficient solutions, machine learning classification emerged as a pivotal approach. Various algorithms, such as probabilistic algorithms and vector machines, played a crucial role in identifying fake URLs. In our project, the focus shifted towards leveraging tree-based machine learning algorithms, including Random Forest, Light GBM, and XG Boost.

These tree-based algorithms have proven to be more effective in addressing the challenges posed by the increasing complexity of malicious URL patterns. To validate their efficacy, we conducted a comprehensive comparison of results, emphasizing the advantages offered by Random Forest, Light GBM, and XG Boost in enhancing the accuracy and efficiency of malicious URL detection. This comparison forms an integral part of our methodologies, showcasing the evolution from traditional techniques to advanced machine learning algorithms in the realm of URL security.

Many different computer programs, called machine learning algorithms, are available to help tackle the problem of finding bad website links. These algorithms are like tools in a toolbox and are documented in books and articles. They provide a strong set of methods to deal with the challenge of figuring out if a website link is harmful or not. It's like having different tools to solve a problem, and researchers use these tools to make the internet a safer place by detecting bad website links.

In this methodology we using different algorithms like Random Forest and Light GBM, which take URL as a input and produces output as Benign, Phishing, Malicious, Defacement.

Here, URL under goes through various steps which includes data collection, data processing, Future engineering, model building, model prediction. A diverse dataset containing Phishing, Defacement, malicious and benign URLs is collected, encompassing various features that contribute to the analysis. Data preprocessing ensures the dataset is cleaned and properly formatted for machine learning. Feature engineering involves extracting relevant characteristics from the URLs, such as length, structure, and domain information.

Model selection, often involving machine learning algorithms like Random Forest or Light GBM, is based on the dataset's characteristics and the desired computational efficiency. Training the chosen model involves using a portion of the dataset to allow the algorithm to learn patterns associated with malicious behavior. Evaluation metrics such as accuracy and precision assess the model's performance, guiding further fine-tuning of hyperparameters.

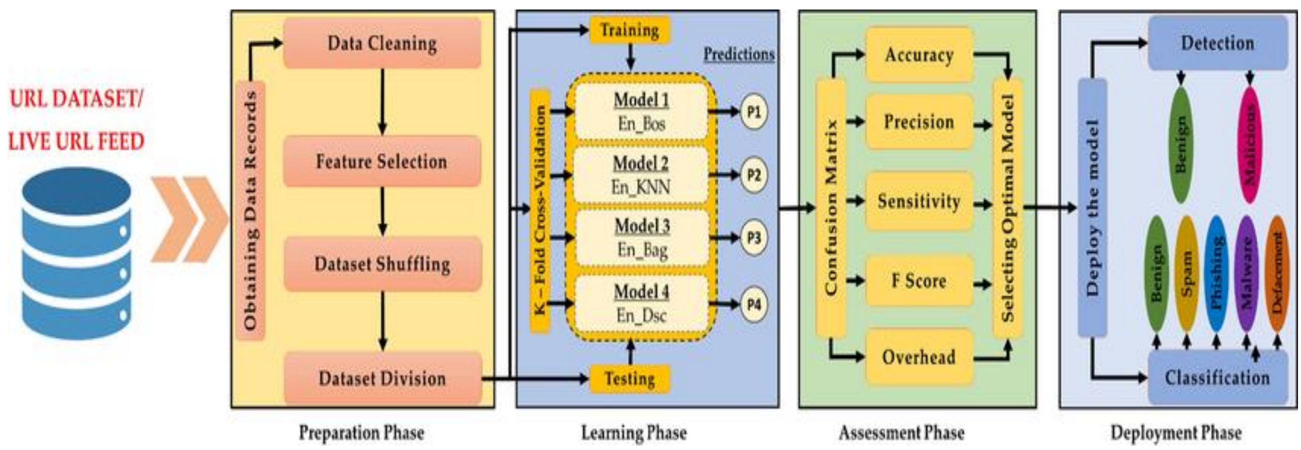


Fig 2: URL patterns classification

3.2 Algorithms:

Validation techniques ensure the model generalizes well to new data, and deployment in real-time or batch settings facilitates actual malicious URL detection. Continuous monitoring and iterative improvement, driven by evolving threats, complete the methodology, creating a robust system for detecting and mitigating malicious URL s. In our study, we employed a comprehensive approach to malicious URL detection by utilizing three state-of-the-art tree-based machine learning algorithms: Random Forest, Light GBM, and XG Boost.

These algorithms were selected for their proven effectiveness in classification tasks above figure shows their ability to handle the complex and dynamic nature of malicious URL patterns. Random Forest, known for its ensemble learning capabilities, and Light GBM and XG Boost, both gradient boosting frameworks, collectively formed a robust framework for accurately distinguishing between legitimate and malicious URLs. The comparative analysis of these algorithms in our research serves to highlight their respective strengths and contributions to enhancing the overall efficacy of our malicious URL detection system

3.2.1 Random Forest:

Random Forest is a popular machine learning algorithm that operates by constructing multiple decision trees during the training phase and outputs the mode of the classes or the average prediction of the individual trees. Each tree in the Random Forest is built independently by selecting a random subset of features and a random subset of the training data. This randomness helps to create diverse and less correlated trees, making the model more robust and less prone to overfitting. During the prediction phase, the Random Forest combines the results from each tree to provide a more accurate and reliable prediction.

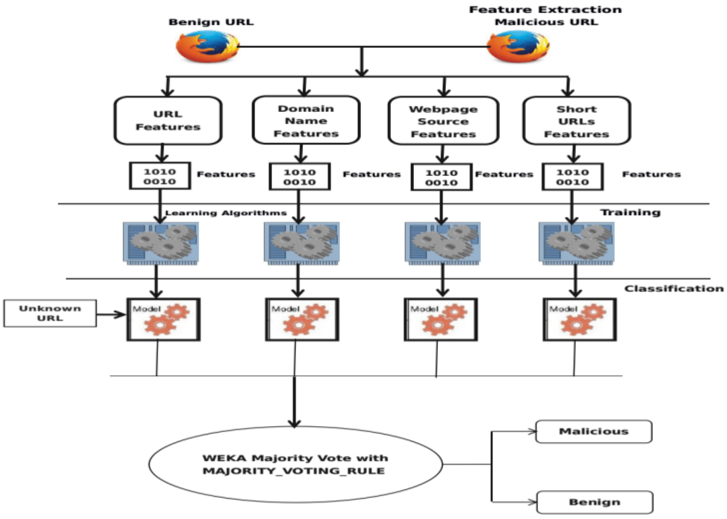


Fig -3 Random Forest classifier

3.2.2 Light GBM:

Light GBM is like a smart tool that helps us find bad website links, especially when it comes to spotting harmful URLs. It's not just any tool; it's really good at its job and works fast. Imagine it as a detective that learns from its mistakes and becomes better at finding bad websites over time. Light GBM is great with big sets of information and lots of different clues (features) that help us tell if a website is safe or not. It's like having a superhero for finding bad online stuff quickly and accurately, making the internet a safer place for everyone

3.3 Data set Description:

The dataset for malicious URL detection comprises a comprehensive collection of 6,51,200 URLs, each associated with specific output labels denoting their nature. The primary objective of this dataset is to facilitate the development and evaluation of machine learning models for the classification of URLs into distinct categories, including benign, defacement, malicious, and phishing.

3.4 URLs (Features): The dataset is rich with URLs, serving as the primary input features for machine learning models. Which consists of **lexical features** from raw URLs:

- **Hostname length:** The length of the hostname
- **Count Dir:** The presence of multiple directories in the URL generally indicates suspicious websites.
- **Count www:** Generally, most of the safe websites have one www in its URL.
- **Abnormal _url:** This feature can be extracted from the database. For a legitimate website, identity is typically part of its URL.
- **Count letters:** The number of letters in the URL also plays a significant role in identifying malicious URLs.
- **Count:** Count finds the count of various special characters like. / , - depending on these count also we can find the URL category
- **Count http:** Generally malicious URLs do not use HTTPS protocols as it generally requires user credentials and ensures that the website is safe for transactions.

3.5 Output Labels (Target): Each URL is associated with specific output labels denoting its classification into one of the following categories:

- **Benign:** Safe and non-malicious URLs.
- **Defacement:** URLs linked to web defacement, indicating unauthorized alterations to the appearance of a website.
- **Malicious:** URLs associated with harmful activities, such as malware distribution or other cyber threats.

- **Phishing:** URLs designed to deceive users into disclosing sensitive information, often by posing as trustworthy entities.

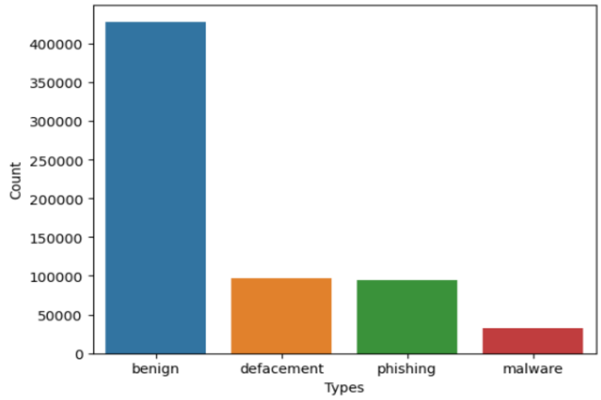


Fig:4 Dataset Labels and count

3.6 Sample Data Set:

URL's	TYPE
br-icloud.com.br	Phishing
mp3raid.com/music/krizz_kaliko.html	Benign
bopsecrets.org/rexroth/cr/1.htm	Benign
espn.go.com/nba/player/_/id/3457/brandon-rush	Benign
http://www.vnic.co/khach-hang.html	Defacement
192.com/atoz/people/oakley/patrick/	Phishing
casamanana.org/education/blba/	benign

The choice of dataset is paramount when embarking on the task of detecting malicious URLs, encompassing threats ranging from malware dissemination to phishing attempts. A comprehensive dataset serves as the cornerstone for training and evaluating machine learning models tasked with discerning between benign and nefarious web addresses.
These datasets typically comprise a diverse array of URLs, meticulously curated to encompass a spectrum of malicious activities prevalent across the cyber landscape. Among the most prevalent threats are URLs associated with malware dissemination, where unsuspecting users are lured into clicking links that trigger the download of malicious software onto their devices. Such datasets often include URLs linked to known malware distribution networks, as well as those generated through automated malware injection techniques. By leveraging such datasets, cybersecurity researchers and practitioners can develop and validate machine learning models capable of effectively identifying and mitigating the risks posed by malicious URLs.

IV. RESULTS AND DISCUSSION

4.1 Results of Descriptive Statics of Study Variables

The results of malicious URL detection signify a critical aspect of cybersecurity efforts. By employing advanced algorithms and machine learning techniques, such as deep learning models and feature engineering methodologies, these systems can accurately identify URLs that pose potential threats. Through extensive analysis of URL characteristics, including domain reputation, URL structure, content analysis, and behavioral patterns, these detection systems can swiftly distinguish between benign and malicious URLs. These results empower cybersecurity professionals to proactively defend against cyber threats, safeguarding users and organizations from phishing attacks, malware infections, and other malicious activities perpetrated through deceptive URLs.

Data collection: A Open Data labeled consisting of 6,51,200 Websites.

	url	type
0	br-icloud.com.br	phishing
1	mp3raid.com/music/krizz_kaliko.html	benign
2	bopsecrets.org/rexroth/cr/1.htm	benign
3	http://www.garage-pirenne.be/index.php?option=...	defacement
4	http://adventure-nicaragua.net/index.php?optio...	defacement

Fig 5- Data set collected

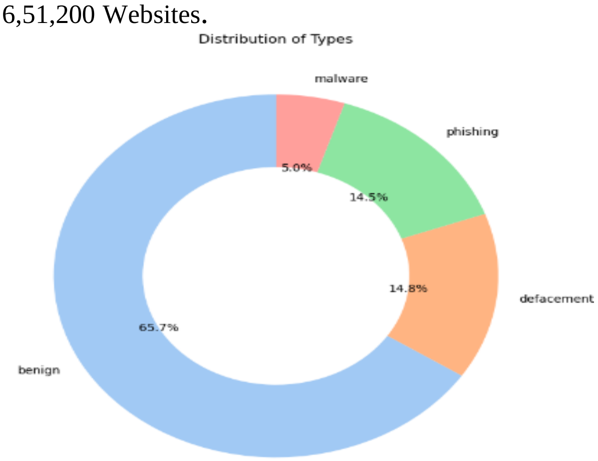


Fig-6 Data set Analysis

Model Evaluation and Comparison:

After fitting the model, as shown above, we have made predictions on the test set. The performance of Light GBM, XG Boost, and Random Forest are shown below.

	precision	recall	f1-score	support
benign	0.97	0.98	0.98	85621
defacement	0.98	0.99	0.99	19292
phishing	0.99	0.95	0.96	6504
malware	0.91	0.86	0.88	18822
accuracy			0.97	130239
macro avg	0.96	0.95	0.95	130239
weighted avg	0.97	0.97	0.97	130239

accuracy: 0.996

Fig -7 Random Forest

	precision	recall	f1-score	support
benign	0.97	0.99	0.98	85621
defacement	0.97	0.99	0.98	19292
phishing	0.98	0.91	0.94	6504
malware	0.91	0.83	0.87	18822
accuracy			0.96	130239
macro avg	0.96	0.93	0.94	130239
weighted avg	0.96	0.96	0.96	130239

accuracy: 0.962

Fig-8 XG Boost

	precision	recall	f1-score	support
benign	0.97	0.99	0.98	85621
defacement	0.96	0.99	0.98	19292
phishing	0.97	0.90	0.93	6504
malware	0.90	0.83	0.86	18822
accuracy			0.96	130239
macro avg	0.95	0.93	0.94	130239
weighted avg	0.96	0.96	0.96	130239

accuracy: 0.959

Fig-9 Light GBM

However, it's essential to interpret accuracy results in conjunction with other performance metrics to gain a comprehensive understanding of the model's strengths and limitations. Metrics such as precision, recall, and F1 score provide additional insights into the model's performance, particularly in scenarios where class imbalances or misclassification costs are prevalent.

Overall, while accuracy results serve as a foundational measure of a Random Forest model's performance in malicious URL detection, they are most informative when considered alongside other relevant metrics, providing a nuanced assessment of the model's efficacy in real-world cybersecurity applications.

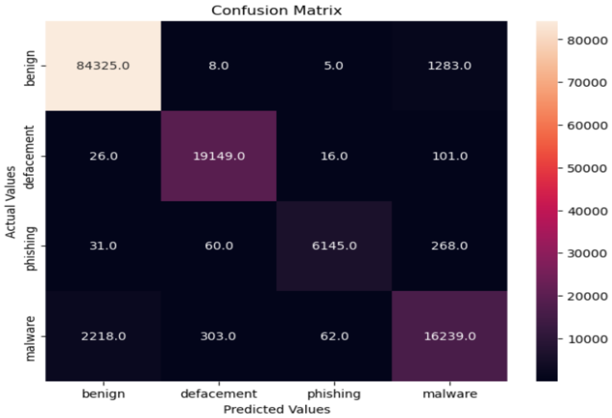
	Predicted Benign	Predicted Phishing	Predicted Defacement	Predicted Malicious
Actual benign	TN	FP	FP	FP
Actual Phishing	FN	TP	FN	FN
Actual Defacement	FN	FN	TP	FN
Actual Malicious	FN	FN	FN	TP

Table: confusion matrix (4 values)

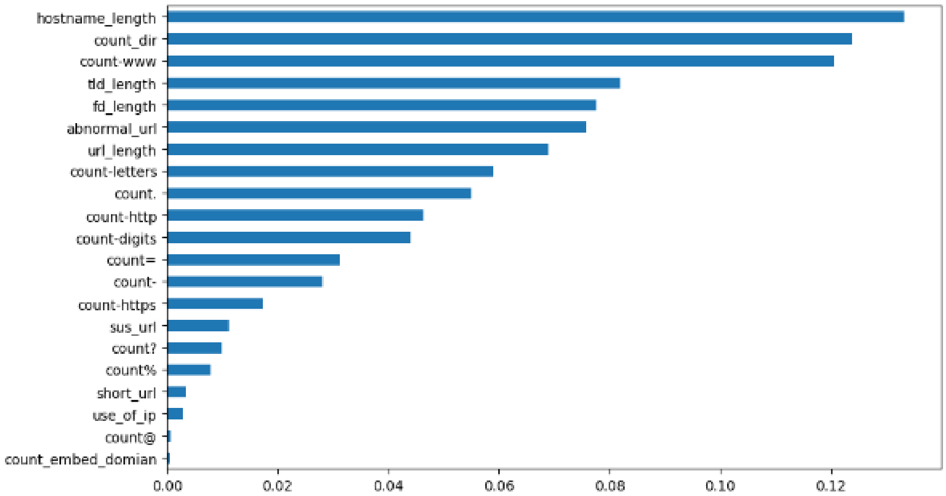
A confusion matrix is a valuable tool for evaluating the performance of a machine learning model.

In the comparison of results from detecting malicious URLs, it has been expressed that the Random Forest algorithm exhibits higher accuracy compared to other detection methods. Random Forest, a popular ensemble learning technique, demonstrates robust performance in distinguishing between malicious and benign URLs by leveraging the collective decision-making of multiple decision trees. Through its ability to handle large datasets with high-dimensional feature spaces and mitigate overfitting, Random Forest effectively captures complex relationships within URL attributes, such as domain reputation, URL structure, and content characteristics. Its ensemble nature enables it to mitigate bias and variance issues commonly encountered in single decision tree models, resulting in improved generalization performance.

The higher accuracy achieved by Random Forest underscores its efficacy in identifying malicious URLs with precision and reliability, making it a preferred choice in many cybersecurity applications. This comparative analysis highlights the strength of Random Forest as a powerful tool in the arsenal of cybersecurity professionals for proactive threat detection and mitigation strategies.



In the analysis of our model's performance, we observed that certain features exhibit notably high precision in predicting specific classes. After training the URLs with the dataset and the machine learning models, when we pass a new URL the machine will predict the output as Malware, Phishing, Defacement, or Benign URL When compared to remaining to models of the Machine learning project Random Forest produces an accuracy of 99 %.




```
urls = ['titaniumcorporate.co.za','en.wikipedia.org/wiki/North_Dakota']

for url in urls:
     print(get_prediction_from_url(url))
```

```
[LightGBM] [Warning] Unknown parameter: silent
MALWARE
[LightGBM] [Warning] Unknown parameter: silent
SAFE
```

5) CONCLUSION:

After subjecting our dataset to rigorous training and testing, particularly against a new set of URLs, our machine learning models underwent comprehensive evaluation. Remarkably, the Random Forest model distinguished itself as a top-performing entity, showcasing an accuracy that seamlessly met our predetermined expectations. The precision and dependability demonstrated by the Random Forest model in effectively classifying the new URLs underscore its resilience within our URL detection system. This thorough analysis not only validates the model's efficacy but also enhances our ability to predict new phishing URLs by leveraging the discerning features embedded within the URL structure. The success of the Random Forest model further solidifies its role as a key component in fortifying our system against emerging threats in the ever-evolving landscape of malicious URLs.

I. ACKNOWLEDGMENT

First, we would like to thank Head of the Department Dr. U. D. Prasan for his extreme support and guidance to accomplish this project. We would also like to thank Sri. Chandra Mohan for his support and advices to get the project completed.

We would like to thank project coordinator Dr. K. Eswara Rao for his extreme support and advices.

We are highly obliged to Director Prof V.V. Nageswara Rao and Principal Dr A.S. Srinivasa Rao, for the facilities provided to accomplish this project.

We are extremely grateful to department staff members and friends who helped us in successful completion of project.

REFERENCES

[1] **“**Malicious URLs detection**”** Shantanu Maheshwari using Machine learning algorithm’s This produces a Higher accuracy of KNN as the 94% ,2021.
[2] **“**Machine Learning for Malicious URL’s Detection**”** Gold Wejinya and Sajal Bhatia in ICT Systems and sustainability by using SVM (support vector machine), Navie Bayes, Logistic Regression Methods ,2021.
[3] **“**Detecting of Phishing URLS on Arabic and English Websites using machine learning” Warmn Ahmed and Noor Ghazi M. Jameel by using a lexical features-based selection, it gives detecting of malicious URLs using Algorithm’s considering Arabic and non-Arabic contents ,2022.
[4] **“**Detecting Malicious URL’s**”** K. Rohit Verma. Manmohana K, Immadisetti Naga Venkata Durga Naveen using the Lexical Analysis Features in addition to 3rd party features, it detecting the malicious classifiers,2020
[5] **“**Malicious URL Detection using Supervised machine Learning Techniques**”** Vara Vundavalli, Farhat Lamia Barsha, Mohammad Masum, Hossain Shahriar, Hisham Haddad using Logistic Regression, Naïve Bayes, and Neural Networks techniques in international Conference on Security of Information and Networks,2020.
[6] **“**Malicious URL Detection Based on Associative Classification**”** Sandra Kumi, ChaeHo Lim and Sang-Gon Lee using CBA (Classification Based on Association) algorithm in entropy MDPI ,2021.
[7] **“**Malicious URL Detection using Machine learning and Deep Learning**”** Amogh, Rajendra, Deekshith, S Parikshith, Asst. Prof. Suma using CNN (Convolutional Neuron Networks) &RNN (Recurrent Neuron Network) in International Research Journal of Modernization in Engineering Technology and Science,2022.
[8] **“**Detecting Malicious URL**”** Bc. Andrea Turiaková using (Support Vector Machine) and TensorFlow in Masaryk University Faculty of Informatics ,2019.
[9] **“**Malicious URL Detection based on Machine Learning**”,** Cho Do Xuan, Hoa Dinh Nguyen, Tisenko Victor Nikolaevich using feature extraction, feature selection in International Journal of Advanced Computer Science and Applications Vol. 11, No. 1, 2020.
[10] **“**A Malicious URL Detection Model Based on Convolutional Neural Network**”,** Zhiqiang Wang, Xiaorui Ren, Shuhao Li, Bingyan Wang, Jianyi Zhang and Tao Yang by using design a malicious URL detection model based on a dynamic convolution neural network (DCNN) in Hindawi Security and Communication Networks, Volume 2021, Article ID 5518528.
[11] **“**Malicious Website Detection using Machine Learning**”,** Mahek Khera, Tanishka Prasad, Liudmila Swati Xess , Rupender Singh, Manpreet Kaur Aiden, by using machine learning model in International Journal Of Engineering Research & Technology (IJERT),volume 11, issue on (05 may 2022).
[12] **“**Malicious URL Detection Using Machine Learning**”,** N. Venkatesh, V. Tejaswini, G. Soumya,T. Shiva Priya , by using blacklist, rule-based detection, anomaly-based detection in Turkish Journal of Computer and Mathematics Education (TURCOMAT), Vol.14 No.02 (2023),537-552.
[13] Canali, D., Cova, M., Vigna, G., & Kruegel, C. (2010). Prophiler: A fast filter for the large-scale detection of malicious web pages. In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security.
[14] Perisic, R., Lee, W., & Lanzi, A. (2010). Using an ensemble of one-class SVM classifiers to harden payload-based anomaly detection systems. In International Symposium on Recent Advances in Intrusion Detection.