



Phishing URL Detection Using Supervised Machine Learning Algorithms

¹Usman Gabriel Ugbede, ¹Sunusi Kabir Alaramma, ²Muhammad Abubakar Ibrahim, ³Adamu Adamu Habu

¹Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi, Nigeria.

²Department of Electrical and Electronics Engineering, Kaduna State Polytechnic, Kaduna, Nigeria.

³University of St Andrews, Jack Cole Building, North Haugh, St Andrews, KY16 9SX, UK.

ABSTRACT

Phishing is one of the security threats that has been in existence for the past decades. It is the process of tricking unsuspecting users with the aid of baits such as URL (Uniform Resource Locator) links to lure victims from a secured platform to an unsecure platform with the intent of stealing valuable information such as credit card details, bank token, and identity etc. Many researchers have developed several methods to combat these threats although been effective, it has failed to combat the increasing adaptive ability of the phishers who attack unsuspecting victims in recent times. In this study therefore, we propose an adaptive data collection system for use in the automatic detection of phishing URL. An adaptive data collection system for legitimate and phishing URL based on certain selected features will be used in the extraction of features using a structured feature extraction vector model which will be trained and tested using five (5) classification models built from Random Forest, Decision Tree, XGboost, Adaboost and KNN. The results obtained from the five models show that model from Random Forest outperforms the other models from other classifiers with an accuracy of 99.38%, precision of 90.21% and a recall rate of 82.55%. The model developed from the random forest classifier was then selected and deployed on a phishing detection platform developed. In the future we seek to expand the features and also seek to explore the usage of deep neural network as the features keeps expanding.

ARTICLE INFO

Article History

Received: March, 2023

Received in revised form: April, 2023

Accepted: May, 2023

Published online: June, 2023

KEYWORDS

Phishing, URL Detection, Machine Learning, Algorithm

INTRODUCTION

Phishing is a criminal mechanism employing both social engineering and technical tricks to steal consumers' personal identity data and financial account credentials. Social engineering schemes use spoofed e-mails, purporting to be from legitimate businesses and agencies, designed to lead consumers to counterfeit websites that trick recipients into divulging financial data such as usernames and passwords. The phishers who want to obtain sensitive data, first create unauthorized replicas of a real website and in recent times, they went as far as impersonating legitimate institutions by sending unsolicited emails using spamming to take individuals from a secured environment to unsecured environments usually from a financial institution or another company that deals with

financial information (Jain et.al, 2018). The e-mail sent is usually created using logos and slogans of a legitimate company to convince the recipient of its legitimacy. The nature and format of Hypertext Mark-up Language makes it very easy to copy images or even an entire website through what is known as cloning of webpages of websites. While this ease of website creation is one of the reasons that the Internet has grown so rapidly as a communication medium, it also permits the abuse of trademarks, trade names, and other corporate identifiers upon which consumers have come to rely as mechanisms for authentication. Phishers then send the "spoofed" e-mails to as many people as possible in an attempt to lure them into the scheme. When these e-mails are opened or when a link in the mail is clicked, the consumers are redirected to a phishing website, appearing to

*Corresponding author: Sunusi Kabir Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



be from the legitimate entity (Baykara et.al, 2022). Online services such as shopping and online banking have brought us a great convenience yet at the same time new threats, like such as giving out confidential information without the knowledge of its users. Phishing has been a serious threat to online security as a lot of people have lost their valuables such as money, landed property because of phishing activity which is usually done using social engineering tactics.

Furthermore, phishing has gone advanced as the physical methods of finding a phishing URL has become difficult as many phishers have gone to the extreme of cloaking the activity of their phishing activities by making it look like a secured environment based on the protocol in which this URL are hosted and this have created a room for concern among many authors and researchers in the last decades.

Anti-Phishing Working Group (APWG) is an organization that collects, analyses, and exchanges a list of credential URLs. It publishes a quarterly report on phishing activities across the globe. The number of phishing attacks has increased from 2021 to 2022 which is the worse ever recorded in its quarterly reports. In its third quarter reports it was observed that more than 1,270,000 reported phishing attacks. Webmail continues to be among the top methods of phishing attempts which grew by 59%. There has been an increase in phishing attacks on named brands which grew to 23.2% and advance fee fraud scams launched by emails increased by 1000% which has become worrisome (APWG, 2022). There are federal laws that punish fraudulent activity in countries around the world where some are functional and others are not because victims of this kind of heinous crimes do not report to relevant authorities as the authority do not have the necessary infrastructure to track the individual because of the highly sophisticated approach that phishers use in defrauding their victims hereby allowing the culprit to get away with the crime which has been the reasons for the increase in the activity of phishing activities around the globe today (Bahnsen et al., 2018).

Jain et.al (2018) proposed a URL-based anti-phishing machine learning method.

They have taken 14 features of the URL to detect the website as a malicious or legitimate to test the efficiency of their method. More than 33,000 phishing and valid URLs were sourced. Support Vector Machine (SVM) and Naive Bayes (NB) classifiers were used to train the models. The phishing detection method focused on the learning process. They extracted the different features that make phishing websites different from legitimate websites. The outcome of their experiment has an SVM model with over 90% accuracy.

Rishikesh et.al (2018) proposed a solution that utilizes machine learning technique for detection of phishing URLs by extracting and analyzing various features of legitimate and phishing URLs. Decision Tree, Random Forest and Support Vector Machine algorithms were used to detect phishing websites. The result of the set of experiment conducted show that Random Forest has the highest accuracy of 89%.

Bahnsen et al. (2018) suggested a more effective method for real-time phishing URL detection. They argued that there are a lot of anti-phishing methods appearing, but scammers use diverse and dynamic methods to scam victims, so a smart and flexible model was needed to detect a phishing URL. The study also identified that data mining methods can be used to promote an active model that contains basic and non-trivial data that can be backed up from huge datasets using classification algorithms to tag a legitimate URL.

Chidimma et al. (2020) proposed detection of malicious URL webpages using deep learning techniques on analyzing HTML code. They used the Convolutional Neural Network (CNN) to learn semantics dependencies in textual content of HTML. 5000 HTML documents were used for the classification, and they obtained an accuracy 93% with relatively lower recall rate. Their approach was able to detect phishing URLs but lacks adequate features to properly improve on the recall rate and precision.

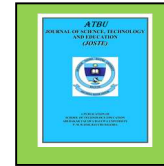
Aljofey et al. (2020) proposed a CNN model to detect a phishing URL. In this study, researchers employed a sequential pattern to capture the URL information. It achieved an

*Corresponding author: Sunusi Kabir Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



accuracy of 98.58%, 95.46%, and 95.22%, respectively on three different benchmark datasets.

Athulya et.al (2020) proposed towards the detection of phishing attacks as applying every technique in a single phase is time-consuming. The study then picked one technique at a given time based on pre-defined condition related to the technique and the dataset used. URL entered will be checked by the Blacklist method or whitelist method or search engine-based techniques. By analyzing all the websites, it was observed that the security feature in the detection website as an administration password for admin to login and able to manage the number of users and read the feedback from the user if there is any feedback relate to login. The user can login with username and password and able to check phishing website and provide feedback related to phishing website. The results of prediction of phishing website using supervised Machine Learning Algorithm recorded the accuracy of 95% for Multinomial NB and 97% of accuracy for Logical Regression.

Dutta (2021) conducted a study that emphasized on the phishing technique used, whereby a phishing URL is seen to include the automatic classification of URLs in a predefined set of category values based on several features and a categorical variable. Machine learning-based phishing techniques rely on URL functions to gather information that can help categorize URLs to detect phishing sites. They argued that to combat the ongoing complexity of phishing attacks and tactics, anti-phishing techniques are essential. The authors used Long Short Term Memory (LSTM) deep learning technique to identify malicious and legitimate URLs. The crawler was developed that crawled 8000 URLs from the Alexa Rank portal, and also used the Phish-tank dataset to measure the efficiency of the proposed phishing URL detector. The result of this study shows that the proposed method provides superior results with better accuracy of 98.3% and F1 within a limited time. A total of 8000 malicious URLs were detected using the proposed URL detector.

Baykara et.al (2022) proposed detection of phishing attacks and spam mails by examining mail contents. Naïve Bayes was used to classify the mails as phishing or legitimate. The study recorded an accuracy of 89%.

METHODOLOGY

In the proposed approach we seek to create a feature extraction vector that will contain the features based on content of the legitimate and phishing URLs into a structured dataset. The collected dataset is then pre-processed and divided into two sets for training and testing the supervised learning algorithm for the study. The data sources are gotten from trancp-list.eu for the legitimate URLs while the phishing URLs were gotten from phishtank.com. The proposed approach is sub divided into the following steps below:

A. Dataset Description

The datasets created are divided into two parts with label 1 for the phishing and 0 for the legitimate URLs. As it was stated earlier the data source features are extracted by the means of scraped data from the URLs provided from the data-source using request and using the beautiful soup features used for the extraction of metadata of an html webpage. This method is shown in Figure 1 below. A feature extraction vector with all the selected content-based features is collected and stored in a structured file.

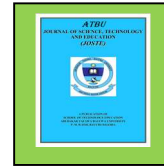
The datasets was built from the data source that has been collected by employing web scraping techniques which involves the sources from phishtank.com and trance-list.eu for the phishing and legitimate URLs. Each content of both sources is collected, extracted based on the selected features and stored into two separate datasets with the labels 1 for phishing URLs and 0 for legitimate URLs. The results of the structured list of the two parts of the dataset as extracted from the data sources are shown in Figures 2 and Figure 3 below with the labels for the phishing and the legitimate URLs.

*Corresponding author: Sunusi Kabir Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



```
def create_vector(soup):  
    return [  
        fe.has_title(soup),  
        fe.has_input(soup),  
        fe.has_button(soup),  
        fe.has_image(soup),  
        fe.has_submit(soup),  
        fe.has_link(soup),  
        fe.has_password(soup),  
        fe.has_email(soup),  
        fe.has_hidden(soup),  
        fe.has_audio(soup),  
        fe.has_video(soup),  
        fe.number_of_inputs(soup),  
    ]  
create_vector()
```

Figure 1: Features extraction by web scraping.

	has_foote	has_form	has_text	has_ifram	has_text	number	ch	has_nav	has_obje	has_pictu	number	c	number	c	number	c	URL	label	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.qfuhb.top/jp.php	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.qsnbpy.top/jp.php	1
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.trea.top/jp.php	1
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.ucwxyw.top/jp.php	1
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.vcuu.top/jp.php	1
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.vnbncj.top/jp.php	1
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.zomoen.top/jp.php	1
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.zttxj.top/jp.php	1
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.zynbuz.top/jp.php	1
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.ceqynl.top/jp.php	1
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.cyrzrh.top/jp.php	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.fakuk.top/jp.php	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.ffreon.top/jp.php	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.fyvku.top/jp.php	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.gbojkt.top/jp.php	1
17	1	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	http://loteriada-urodzinowa.pl/millenium_payment.php	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.hlrooh.top/jp.php	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.hyhowu.top/jp.php	1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.jyxouz.top/jp.php	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	http://www.eki-net.con-aesccceesaas.mtbuiy.top/jp.php	1

Figure 2: Results of Structured data set for the phishing URLs

*Corresponding author: Sunusi Kabir Alaramma
✉ skalarammane@gmail.com
Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,
© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved

	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT
1	number_c	number_c	number_c	has_foote	has_form	has_text	has_ifram	has_text	number_c	has_nav	has_objec	has_pictu	number_c	number_c	number_c	URL	label	
2	1	13	0	0	1	0	0	0	2	0	0	0	0	8	1	http://google.com	0	
3	0	292	0	0	0	0	1	1	7	0	0	0	0	1	0	http://youtube.com	0	
4	2	32	0	0	1	0	0	1	18	0	0	0	0	2	0	http://facebook.com	0	
5	17	184	0	1	1	0	1	0	26	1	0	1	44	27	0	http://microsoft.com	0	
6	6	84	0	0	1	0	0	0	20	0	0	0	2	31	0	http://netflix.com	0	
7	57	1057	30	1	1	0	0	1	23	1	0	0	0	104	0	http://amazonaws.com	0	
8	1	8	0	0	1	0	0	0	12	0	0	0	0	1	0	http://twitter.com	0	
9	2	2	0	0	0	0	0	0	19	0	0	0	0	1	0	http://instagram.com	0	
10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	http://baidu.com	0	
11	0	100	9	1	1	0	0	1	16	1	0	0	0	68	0	http://apple.com	0	
12	6	53	0	1	1	0	0	1	20	1	0	0	0	86	0	http://linkedin.com	0	
13	1	86	0	0	1	0	0	0	3	0	0	0	0	58	0	http://wikipedia.org	0	
14	3	349	0	0	0	0	0	0	16	1	0	0	0	92	0	http://cloudflare.com	0	
15	57	961	0	0	1	0	0	1	14	0	0	0	0	54	0	http://yahoo.com	0	
16	27	335	0	0	1	0	0	1	9	0	0	0	0	75	0	http://qq.com	0	
17	7	139	3	1	0	0	1	0	22	1	0	0	0	32	0	http://live.com	0	
18	31	273	0	0	0	0	0	0	11	0	0	0	0	98	0	http://bilibili.com	0	
19	277	489	0	1	1	0	1	1	31	1	0	1	75	140	0	http://azure.com	0	
20	31	141	0	1	0	0	1	0	14	1	0	0	0	87	0	http://fastly.net	0	
21	11	71	12	1	1	0	1	0	17	1	0	0	0	53	0	http://wordpress.org	0	

Figure 3: Results of Structured data set for the legitimate URLs

The dataset was created to be a balanced dataset as the combined list is made up of 26,587 websites URLs of which comprises of 16062 legitimate URLs from tranco-list.eu and 10525 phishing URLs from phishtank.com which

makes up 39% for the phishing URLs and 61% for the legitimates URLs hence a balanced dataset is been used. The results for the balanced datasets is shown in the Figure 4 below.

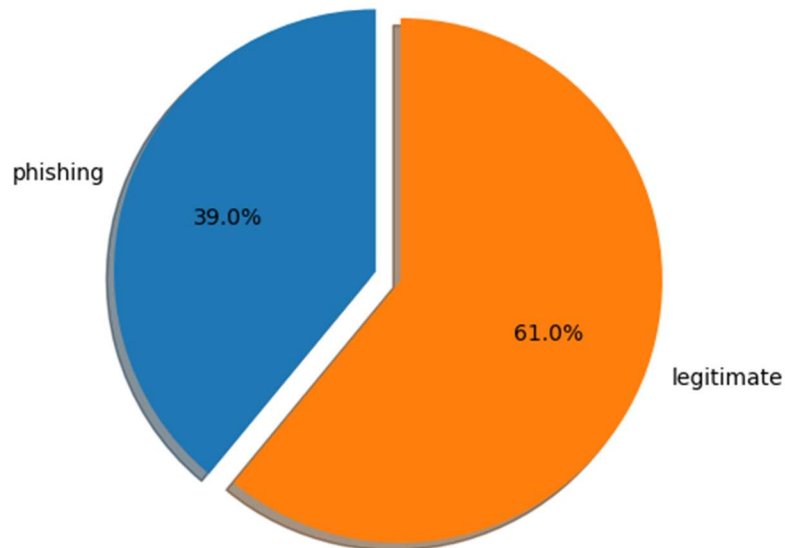


Figure 4: Results for the combined structured dataset

*Corresponding author: Sunusi Kabir Alaramma
 ✉ skalarammane@gmail.com
 Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,
 © 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved

D. Model Used

Five (5) models were built from five different supervised machine learning algorithms as used in this study. The classifiers used are Random Forest, K-Nearest Neighbor, Adaptive Boosting, Decision Tree and XGboosting. For each of the classifiers, portion of the dataset was used for training and the other portion for testing the performance of the various models.

E. Confusion Matrix

The confusion matrix is a table that shows the classification of the case where a model correctly predicts a phishing URL which is denoted as True positive (TP), while a case where a URL is wrongly classified as legitimate is denoted as True negative (TN). There is also a case where a URL is classified as phishing is denoted as False positive (FP). Lastly a case where a URL is wrongly classified as a legitimate URL but is phishing in real sense is denoted as False Negative (FN). The Table 1 below shows how the confusion matrix is evaluated.

Table 1: confusion matrix evaluation

	Predicted legitimate URLs	Predicted phishing URLs
Actual legitimate URLs	TP	FN
Actual Phishing URLs	FP	TN

F. Performance Metrics

In the performance evaluation the performance metrics used for the learning algorithms are shown in the equation below for the accuracy, precision and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \text{ ----- (4)}$$

$$\text{Precision} = \frac{TP}{TP + FP} \text{ ----- (5)}$$

$$\text{Recall} = \frac{TP}{TP + FN} \text{ ----- (6)}$$

RESULTS AND DISCUSSION

The results for running the algorithms to create a model from the datasets are shown based on the classification algorithm that was used and the selected features that the classification algorithm is fed with. The result is shown in the Figure 6 below.

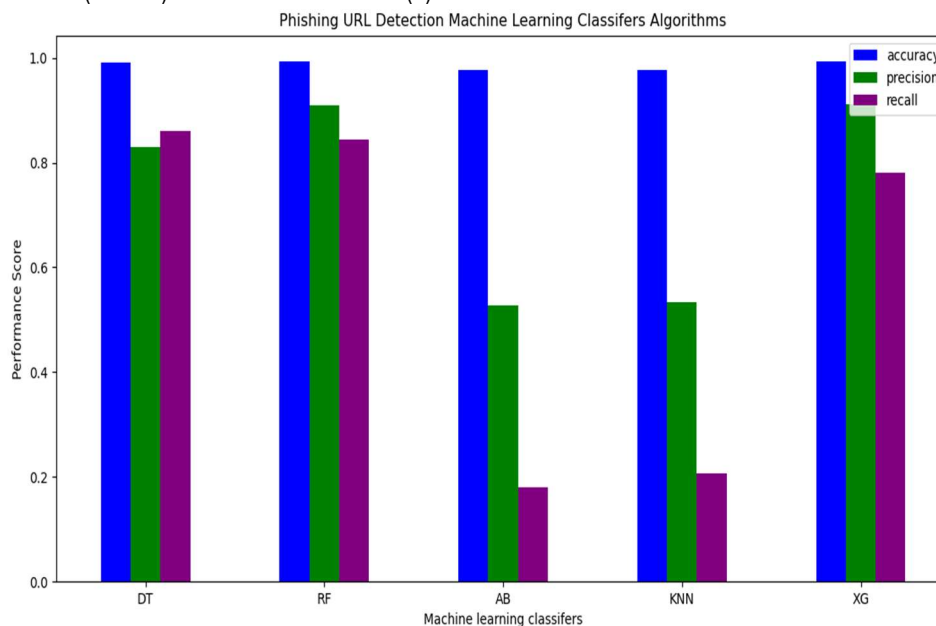


Figure 6: Results for phishing URL detection

*Corresponding author: Sunusi Kabir Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



The result for the performance evaluation is shown in Table 2 below for the

classification algorithm based on the accuracy, precision and recall.

Table 2: Comparison of Performance Evaluation

Classification Algorithm	Accuracy	Precision	Recall
Decision Tree	0.9918	0.8372	0.8541
Random Forest	0.9938	0.9021	0.8255
AdaBoost	0.9770	0.5243	0.1781
K-Nearest Neighbor	0.9770	0.5288	0.1953
XGBoost	0.9928	0.9103	0.7819

DISCUSSION OF RESULTS

From Table 2 the performance metrics for the accuracy of the classifiers used shows that Decision Tree has 99.18% accuracy based on the level of classifying the phishing URLs from the legitimate URLs. In this model, 24 URLs were predicted to be actual legitimate URL, while 1481 URLs were classified to be actual Phishing URL. However, 1 URL was predicted to be phishing instead of been a legitimate URL, also 1 URL was labeled legitimate whereas it is a Phishing URL on the dataset. Random Forest shows 99.38% accuracy based on the level of classifying the phishing URLs from the legitimate URLs. This result was achieved because 24 URLs were predicted to be actual legitimate URLs, while 1481 URLs were classified to be actual Phishing URL. Nevertheless, 2 URLs was predicted to be phishing instead of been a legitimate URL, also 5 URLs was labeled legitimate whereas it's a Phishing URL on the training set. Adaboost recorded an accuracy of 97.70% based on the level of classifying the phishing URLs from the legitimate URLs this is because, 9 URLs were predicted to be actual legitimate URLs, while 1467 URLs were classified to be actual Phishing URL. But 23 URLs were predicted to be phishing instead of been legitimate URLs, and 8 URLs were labeled legitimate whereas they are Phishing URL on the dataset. K- Nearest Neighbor obtained an accuracy of 97.70% based on the level of classifying the phishing URLs from the legitimate URLs. However, from the model, 2 URLs were predicted to be actual legitimate URL, while 1448 URLs were classified to be actual Phishing URL. But, 51 URL were predicted to be phishing instead of been legitimate URLs, also 6 URLs were

labeled legitimate whereas they are Phishing URLs on the dataset. XGBoost recorded 98.28% accuracy based on the level of classifying the phishing URLs from the legitimate URLs this is for the reason that, 11 URLs were predicted to be actual legitimate URLs, while 1448 URLs were classified to be actual Phishing URLs. But, 42 URL was predicted to be phishing instead of been a legitimate URLs, and 6 URLs was labeled legitimate whereas it's a Phishing URLs on the dataset.

However Random Forest outperformed other classification algorithm because of its ability to accurately learn from the actual set to predict the legitimate and phishing URL with minimal error.

In addition, from Table 2 the precisions for the classification algorithms show that Random Forest, Decision Tree and XGboost shows higher precision rate of 90.21% , 83.72% and 91.03% respectively while Adaboost and KNN shows an average precision rate of 52.43% and 52.88% respectively because of their slow learning rate in specifying features labels in the category.

Furthermore, from Table 2 the recall for the classification algorithm indicated that Random Forest, Decision Tree and XGboost recorded higher recall rate of 82.55%, 85.41% and 78.19% while Adaboost and KNN shows a very lower recall rate of 17.81% and 19.53% respectively because of their cost of calculating the distance between the new point and each existing point is huge which degrades the performance of the algorithms.

However, Random Forest has recorded the highest overall performance and hence outperform other classifiers because of its good

*Corresponding author: Sunusi Kabir Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved

prediction capability in specifying in exact the features category based on the label it learns from. However other learning algorithms shows a reasonable precision learning rate as in specifying features categories based on the learning label set.

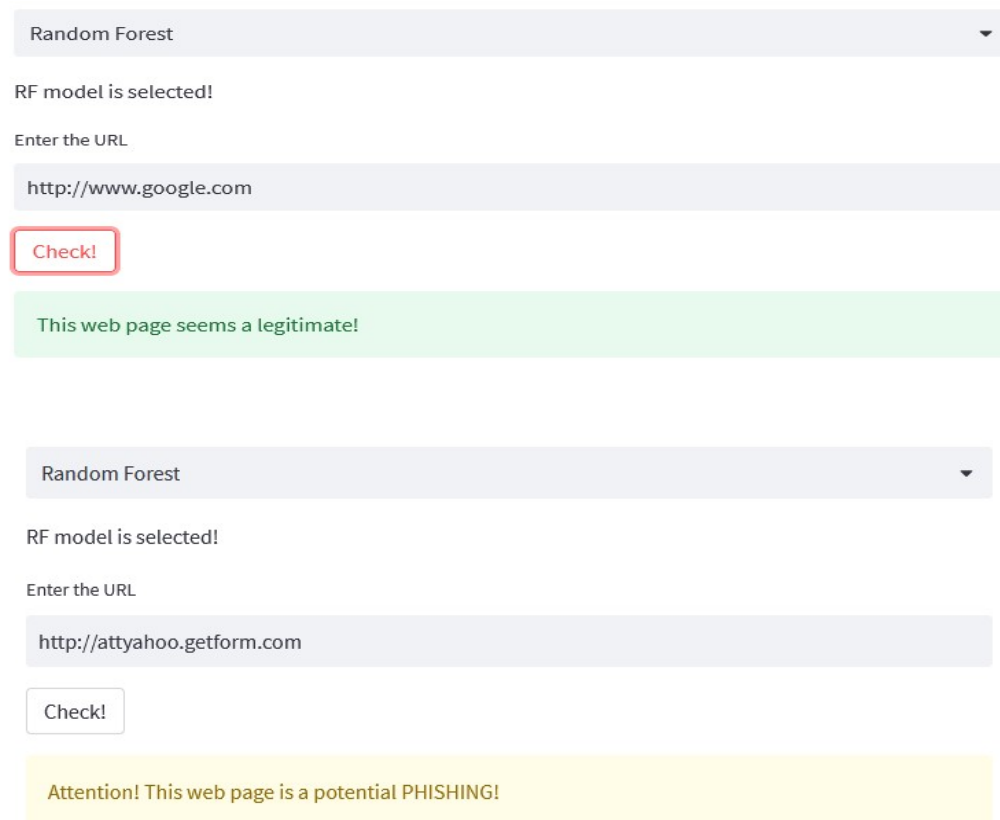
Finally, in the case of Decision Tree still outperform other classifiers in the rate of recall for the reason that of its good prediction capability in detecting features category based on the label it learns from. However other learning algorithms such as Random Forest and XGboost shows a reasonable recall learning rate as in detecting

features categories based on the learning label set.

Model Deployment

Based on the comparison Random Forest was selected to be deployed based on the research work because it shows higher accuracy, precision and recall. Figure 7 above shows the deployment of the best models selected on stream-lit frame work which is used for the testing of the models and to see the result on a graphical user interface in form of a web browser.

The result for the deployment of the selected model is shown in the Figure 7 below:



The screenshot displays two instances of a web application interface. Each instance features a dropdown menu set to 'Random Forest', a text input field for a URL, and a 'Check!' button. The first instance shows a green feedback message: 'This web page seems a legitimate!' for the URL 'http://www.google.com'. The second instance shows a yellow feedback message: 'Attention! This web page is a potential PHISHING!' for the URL 'http://attyahoo.getform.com'.

Figure 7: Deployment Of best Model on Stream-lit framework

CONCLUSION

Machine learning is effective for drastically reducing the threats of phishing URLs in today's world and in the future. The study shows

the effectiveness of supervised learning classifiers in the detection of phishing URLs and also based on the performance, Random Forest Machine Learning algorithms performed better than the

*Corresponding author: Sunusi Kabir Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



other classifiers compared in the course of the study in the classification of phishing URLs even though others still prove useful in the detection of phishing URLs.

REFERENCES

- APWG (2022) phishing activity trends report, 4th quarter 2022, tech rep, pp3-5 retrieved from <http://www.apwg.org/>
- Aljofey, A, Jiang Q, Qu Q, Huang M, Niyigena JP (2020). An effective phishing detection model based on character level convolutional neural network from URL. *Electronics*. Sep; 9(9):1514.
- Athulya, A. A. (2020) Towards the Detection of Phishing Attacks, *4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184)*, Tirunelveli, India pp.337343,doi:10.1109/ICOEI48184.20.9142967.
- Bahnsen, et al. (2018), How to Detect Phishing Website Using Three Model Ensemble Classification, *4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184)*, Tirunelveli, India.
- Chidimma, et al (2020), HTML Phish: Enabling phishing web page detection by applying deep learning techniques on HTML analysis. 19(3) <https://doi.org/1909.01135v3>
- Dutta, K, (2021), Detecting phishing websites using machine learning technique. 16 (10): e0258361 16(10). <https://doi.org/10.1371/journal.pone.0258361>
- Jain, A. K. and Gupta, B. B. (2018), PHISH-SAFE: URL Features-Based Phishing Detection System Using Machine Learning, *Cyber Security, Advances in Intelligent Systems and Computing*, vol. 729, Doi: 10.1007/978-981-108536-9_44.
- Janson, K, & Von S. R. (2011), Phishing for phishing awareness. 32(10) <https://doi.org/10.1080/0144929X.2011.632650>.
- Rishikesh, Mahajan and Irfan Siddavatam (2018) "Phishing website detection using machine learning" *International Journal of Computer Applications (0975 – 8887) Volume 181*.

*Corresponding author: Sunusi Kabir Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi,
© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved