EVALUATION OF PERFORMANCE MEASURES FOR RELIABLE AND SECURE PHISHING DETECTION SYSTEM

Pratikkumar A Barot[1*], Sunil A Patel[2], H B Jethva[3]

•

Department of Computer Engineering, Gujarat Technological University, Ahmedabad, India

[1]pratikabarot@gmail.com

[2]dp.sapatel@gmail.com

[3]hbjethva@gmail.com

Abstract

Phishing is an illegal act and security breach which acquires a user's confidential information without consent. Anti-phishing techniques used to detect and prevent such malicious acts to provide data safety to the end user. Researchers proposed an anti-phishing solution with the help of techniques like the blacklist record, heuristic function, visual similarity, and machine learning algorithm. In recent times many researchers proposed machine learning techniques for phishing detection and achieve more than 90% accuracy. However, there is reliability issue in the accuracy measures used by the researchers. In real life, the phishing dataset is unbalanced. Most of the researchers ignore this data quality during their research work design. In the case of unbalanced data, traditional accuracy measure does not give proper performance evaluation. It shows biased performance evaluations. In this paper, we experimented with an unbalanced dataset of phishing detection and did detailed result analysis to highlight the reliability issue of traditional performance evaluation measures for unbalanced data classification. We experiment with four classification algorithms and found that more than 90% of accuracy does not entitle any classifier as secure and safe if the dataset is unbalanced. Our work highlights the data factors and algorithmic limitations that compromise the system security and data safety.

Keywords: Reliable phishing detection, Cyber security, Unbalanced data, Performance evaluation, Cyber safety.

1. Introduction

The use of internet-based services increases day by day. However, the internet brings some hidden security threats to users. One of the threats is Phishing. Phishing is a cybercrime that acquires the user's confidential information without the consents of the user. Phishing websites looks like its legitimate counterpart and spoof user activity to obtain user's personal information [1]. Other phishing techniques which compromise the data safety are email phishing, SMS phishing, voice phishing, and website phishing [2].

The Anti-Phishing working group (APWG) [3] is a non-profit organization. APWG examines and publishes reports of phishing attacks. According to the APWG report, in the 1st quarter, 0f 2018 total 2,63,538 phishing attempts gets detected, and in the 2nd quarter of 2018, total 2,33,040 phishing attempts gets detected. The online banking and payment sector experiences the most number of phishing attacks.

Many researchers performed studies to proposed reliable and secure phishing detection tools [4–6]. However, the unbalanced nature of the phishing dataset does not much explored by researchers. The phishing dataset is unbalanced dataset with two class labels: Legitimate and Phishing. It has more phishing instances as compared to legitimate instances. In most of the existing research work, the legitimate class label is in a minority. Unbalanced data classification needs more focus when the misclassification cost of one class is higher than another class [7]. In the phishing dataset, the misclassification cost of phishing URLs is much higher than legitimate URLs. It makes phishing data classification more complex. Further the unbalanced data makes the traditional evaluation parameters to present misleading performance evaluation [8].

In this paper, we performed a detailed study and in-depth result analysis to highlight the shortcomings of performance measurement in existing phishing detection researches. To propose an accurate phishing detection system, identification and validation of the performance evaluation parameter is primary requirement. The biased performance measures present biased evaluation of the security system. Especially, for the unbalanced data, we need a balanced and unbiased performance evaluation technique [7].

This paper starts with the introduction section. The second section discuss the related work of phishing detection system. Third section gives dataset information, and fourth section give detail of implementation methodology for our experiment. Fifth section present the unbiased and reliable performance measures, and compares the results with traditional performance measures to highlight the shortcoming of widely used metrics for reliable evaluation of machine learning based security systems.

2. Related Work

Machine learning is widely use in designing of cyber security system. It is increasingly used for system and data safety through effective fault detection and bug management [9]. Various anti-phishing techniques have been use for the cyber security. Researchers uses techniques such as user training, black-list approach, heuristic-based approach, visual similarity-based approach, and machine learning approach for the anti-phishing. The subsequent section summarises the widely used anti-phishing techniques.

a) User education or simulated training

Many researchers worked on the development of mobile games which helps to educate people and to raise awareness about phishing among the uses [4]. The approach is limited to users who love gaming and due to this; this approach alone is not sufficient for phishing detection [10].

b) Blacklist approach

In this approach, a dataset is created and maintained for a list of phishing URLs. Target URL is checked in this dataset. If found in a phishing dataset then it is detected as phishing otherwise it is called legitimate [5,11].

c) Heuristic-based approach

In this approach, the heuristic function is applied to extracted features or to extract features for phishing detection. However this approach is not able to detect all new attacks and it is easier to bypass once the attacker knows the heuristic algorithm or features used [12].

d) Visual similarity-based approach

The limitation of this is approach is more time is needed for image comparison and more space is needed to store all images [6].

e) Machine learning-based approach

Machine Learning is the most promising technique approach than other anti-phishing solutions [13]. Machine Learning-based classifiers are efficient classifiers and achieved an accuracy of more than 90% [14,15]. The machine learning-based anti-phishing technique is robust and more accurate as compared to other techniques [13,14]. Classification is one of the machine learning techniques for label prediction. Classification algorithms like support vector machine (SVM), Naïve Bayesian, KNN, decision tree, and ensemble classification algorithm like random forest are widely used. Random forest is a collection of more than one decision tree algorithm and used an ensemble technique for final prediction. Many authors claim that the random forest gives better accuracy in phishing website detection [2,15–19].

Jain et al. [1] proposed a URL based machine learning technique for phishing detection. The authors use SVM and achieved 91.28% accuracy. However, in the case of redundant or irrelevant features performance gets decreased. Authors use accuracy for performance evaluation which present majority biased performance evaluation [20].

Alejandro et al. [18] proposed a neural network algorithm for the detection of phishing URLs. They proposed a recurrent neural network (RNN) based technique. RNN requires more time to train the model. Sudhanshu et al. [21] proposed a rule-based classification technique for phishing website detection. They found that the association classification algorithm performs well as compared to other algorithms. Bayu et al. [17] presented a comparison of ensemble approaches like random forest, rotation forest, gradient boosted machine, and extreme gradient boosting against decision tree. The result shows the superior performance of random forest as compared to other classifiers.

Anand Desai et al. [22] uses SVM, KNN, and random forest classifier for malicious web content detection from balance data. Jain et al. [2] proposed an anti-phishing approach that extracts features from the client-side. The proposed approach is fast but it extracts features only from URL and source code. This approach has a limitation as it can detect webpages written in HTML only. Ahmad et al. [23] proposed three new features to improve the accuracy rate for phishing website detection. Liu et al. [19] proposed an approach that focuses on character frequency features. In this, they have combined statistical analysis of URL with a machine learning technique. However, the authors use majority class centric performance evaluation parameters in their experiments. In 2016, Tahir et al. [24] proposed a hybrid model for the classification of the phishing website.

In 2015, Bhagyashree et al. [25] proposed a feature-based approach. Features like WHOIS, Page Rank, Alexa rank, and Phish Tank-based features are used for disguising phishing and non-phishing website.

Most of the researchers claim more than 90% accuracy in phishing detection. However, the phishing dataset is unbalanced. An unbalanced dataset has unequal class distribution. Accurate detection of both the class is important in the case of an unbalanced phishing dataset. As per Barot et al. [20], traditional accuracy measure gives biased result for unbalanced data. So validation and verification of performance evaluation for the phishing detection technique is an important task. In existing researches the 90% accuracy is mostly due to the accurate classification of majority class.

Traditional classification algorithms are designed for balanced data and thus do not perform well for unbalanced data [8]. Unbalanced data has skewed class distribution. In binary-class unbalanced data, one class is in majority and another class is in the minority [26]. Imbalance ratio (IR) is used to measure the level of imbalance and it is derived as the ratio of majority class instances and minority class instances [26]. Unbalanced data need special care to improve classification accuracy [27].

As per Barot et al. [20], an accurate and unbiased evaluation of unbalanced data classification is an important task. Traditional metrics are not suitable for performance evaluation of unbalanced data classification. Especially, when the misclassification cost of minority class is huge as compared to the misclassification of the majority class, we need unbiased metrics for performance evaluation. In the case of the phishing dataset, data is unbalanced and misclassification cost is different for both the minority and majority class. So we need unbiased performance evaluation measures. Barot et al. [20] proposed a balanced and unbiased performance evaluation for unbalanced data classification.

Table 1: *The original Data source of phishing and legitimate URL.*

Dataset	No of instances	Category
PhishTank [28]	30251	Phishing
DMOZ [29]	3494	Legitimate

3. Dataset

In a real-time scenario, legitimate URLs are more than the phishing URLs. It requires a robust phishing detection system that accurately identifies phishing URLs among the large numbers of legitimate URLs. This empirical study is designe to verify the performance evaluation of phishing detection from unbalanced phishing data. We use PhishTank [28] for phishing and DMOZ [29] as a non-phishing dataset. There are a total of 30251 instances of phishing and 3494 instances of legitimate class. We combine these two datasets into one dataset. The resultant dataset is unbalanced in which the phishing class is in majority, and the legitimate is in minority.

Table 1 shows detail of the original data source of phishing and legitimate URLs. For our experiment, we create a pre-labelled unbalanced dataset that comprises extracted features of URLs (label 1 for phishing and 0 for legitimate). Total 11 features namely, "IP Address", "presence of subdomain", "'@' Present in URL or not", "Presence of dash(-) in URL or not", "Length of URL", "Suspicious words in URL", "Embedded Domain", "presence of HTTPS Protocol", "HTTP_Count", "DNS lookup", and "Inconsistent URL" are extracted from the URLs. URL is treated as inconsistent if the domain name does not match in the WHOIS database.

We created a dataset by taking 23000 phishing URLs from PhishTank [28] and 2000 legitimate URLs from DMOZ [29]. The final dataset contains a total of 25000 URL instances. The phishing dataset is unbalanced and legitimate instances are in minority.

4. Implementation Methodology

For our experiment, we have used the Naïve Bayesian, Sequential minimal optimization (SMO), decision tree, and random forest algorithms. The main aim of this experiment is to check algorithms reliability by analyzing the impact of uneven distributions of classes on classification accuracy. We used the WEKA library for the implementation. We had kept the phishing dataset unbalanced to highlight the main reasons of the high accuracy value even if minority class is totally misclassified. The poor performance of minority class detection hidden under the accurate detection of the majority instance makes the security system unreliable.

Initially, phishing and legitimate URLs given to the features extractor that extracts the features. The extracted features then used for the detection of phishing URLs. The extracted features used to train the naïve Bayesian, SMO, random forest, and decision tree classifiers.

4.1. Main Steps of Proposed Approach

The major steps of our experimental study are:

Step 1) Dataset preparation: Dataset is prepared by collecting phishing URLs from PhishTank and legitimate URLs from the DMOZ directory.

Step 2) Data pre-processing: the dataset is pre-processed to remove noise and to fill missing values. The missing value is handled by taking attribute mean.

Step 3) Feature extraction: extraction of features from the website URL by using a feature extraction algorithm.

Step 4) Train model: Traditional algorithms are implemented in JAVA using the WEKA library.

Using training dataset classifiers are trained.

Step 5) Test Model: Trained model is tested with test split and performance is evaluated.

Step 6) Result and evaluation: analysis of result and checking reliability of performance matrices.

We have used 10-fold cross-validation for training and testing. Figure 1 shows the flow of our empirical work.

As shown in Figure 1, first we collect legitimate and phishing URLs. We applied feature extraction and feature selection methods to create dataset for classification algorithms. Then we train the traditional classification model from the training set and test the model using the test set. Finally, we present the results and discuss the importance of the unbiased performance measure for unbalanced dataset classification.

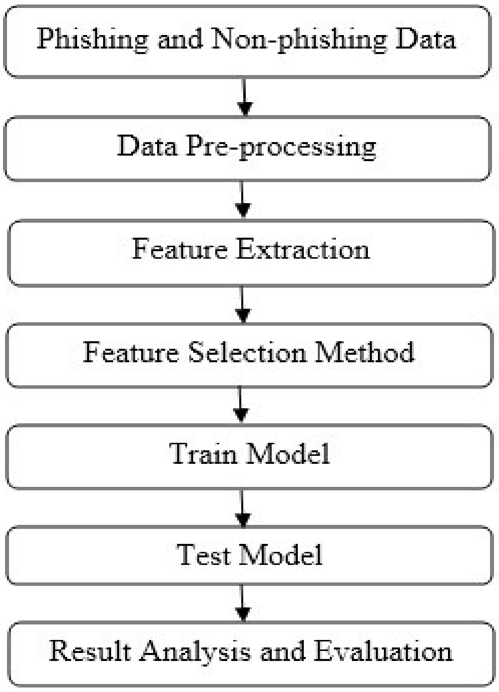

Figure 1: *Implementation flow*

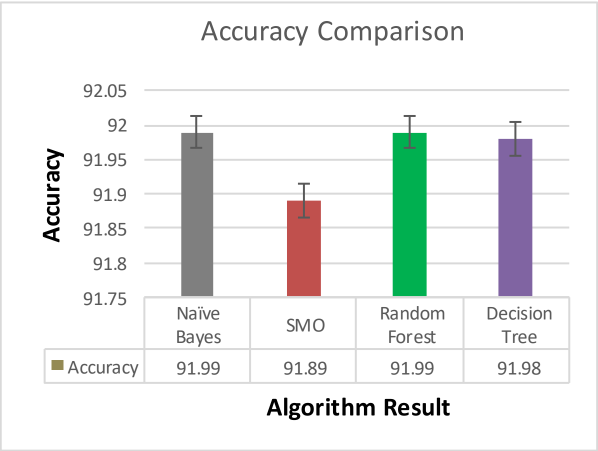

Figure 2: *Performance (Accuracy) Comparison of Classifiers*

5. Result and Discussion

Most of the existing researches show high accuracy of phishing detection. Our experiment also get similar performance in terms of high accuracy. Figure 2 shows the results comparison of naïve Bayesian, SMO, decision tree, and random forest classifiers.

Despite the high accuracy, detailed result analysis reveals that the algorithms are not an optimal choice for phishing detection because of unbalanced nature of the phishing dataset. The detailed results (TP_rate, precision, and F-measure) of naïve Bayesian, random forest and decision tree given in Table 2. The naïve Bayesian, random forest and decision tree shows identical results for phishing detection.

In Table 2, the accuracy, TP rate, precision and f-measure value for the individual class is shown to highlight the shortcoming of the traditional measures. As shown in Table 2, the accuracy, TP rate, precision, and f-measure for the minority class is 0, and for majority class, it is more than 90%, that produce the weighted average close to 90%. If classifiers are selected just by looking into the weighted average value, then it increase misclassification cost due to the misclassification of all legitimate instances. Legitimate instances are small in number (minority class) and thus ignored by all three classifiers.

The confusion matrix in Table 3 shows that one instance of the majority class gets misclassified as minority class, and all legitimate (minority) class instances gets misclassified as phishing class (majority) instances. Algorithms that show a similar pattern of the results should not considered reliable algorithm for phishing detection due to their inability of correct predictions of minority class.

From the result, we observed that the high accuracy of classification algorithms is due to the high predictive performance of the majority class. Classification accuracy of minority class is close to zero due to the high imbalance ratio of the phishing dataset and the biasing of traditional classifiers towards the majority class. Many researchers consider overall accuracy as a performance measure and show more than 90% accuracy to claim good performance of their algorithms. However, the result analysis suggest that majority class biased accuracy is not an appropriate performance measure for the unbalanced data classification [8]. The security system designed using such machine learning algorithms are unreliable and do not gives proper safe environment.

Table 2: *Detailed Result*

	Accuracy	TP Rate	Precision	F-measure
Phishing	0.999	0.999	0.920	0.958
Legitimate	0.000	0.000	0.000	0.000
Weighted Avg.	0.919	0.920	0.846	0.882

Table 3: *Confusion Matrix of Naïve Bayesian.*

	Phishing	Legitimate
Phishing	22999	1
Legitimate	**2000**	**0**

In our experiment, all the traditional classification algorithms claims around 90% for the accuracy, TP rate, precision, and f-measure. However, none of the classifier is optimal because of the tendency of the classifiers to consider all the instances as phishing instances. As the phishing class is in majority and covers 92% of the dataset, if the classifier considers all instances as phishing, then it automatically achieves almost 90% accuracy. However, it hides the poor prediction of legitimate URL.

In the case of class imbalance, such high value of performance measures is misleading. The actual low accuracy of minority class detection outnumbered by the high accuracy of the majority class detection. Such classifiers are unreliable and not provide optimal solution for secure system design. Such machine learning algorithms should not employ even though they gives more than 90% accuracy. Selection of classifier just because it shows more than 90% accuracy ends up by blocking of all the URLs – both legitimate and phishing URLs get blocked as each URL detected as a phishing URL. Such a bias algorithm presents bias results and shows high accuracy because of the dominance of majority class instances.

Barot et al. [20] proposed a new unbiased evaluation parameters for unbalanced data. They proposed two measures called B-mean and IR based weighted mean named IRWMean. The B-mean gives a more balance performance evaluation while IRWMean gives minority class-biased performance evaluation. The mathematical equations for the B-mean and IRWMean are given in Eq. (1) & Eq. (2).

$$IRWMean = (IR \times TN\_rate) + (1/IR \times TP\_rate) \div (IR + 1/IR) \tag{1}$$

$$B\text{-}mean = ((IR{\times}TN\_rate) + (1/IR{\times}TP\_rate)) \div (((IR{+}1/IR) + Acc) \div 2)) \tag{2}$$

Table 4, shows performance evaluation of the naïve Bayesian algorithm in terms of the B-mean and IRWMean for the phishing dataset. As shown in the table, B-mean is 0.46. The value of B-mean is very low as compared to the accuracy, precision and f-measure values given in Table 2. This is because of poor prediction of the minority class. The B-mean consider the imbalanced ratio to determine the misclassification cost of the minority class and majority class. Although the all majority class instances are correctly classified, the B-mean is 0.46 to indicate the misclassification of costly minority class. The B-mean show more balanced performance evaluation while the accuracy, f-measure and precision are majority class biased measures.

The value of IRWMean is 0.016 which indicate total misclassification of costly minority class. The IRWMean considers negligible benefit of correct prediction of majority class. Proper tuning of the weight according the domain specific misclassification cost gives more balanced performance evaluation.

Figure 3 shows a comparison of traditional performance parameters with IRWMean and B-mean. As none of the minority instances is correctly classified the TN_rate (for the legitimate class) is zero. But still, the precision, f-measure, and accuracy are too high and they give biased and misleading results. From the accuracy, precision, or f-measure, we are not able to discover that the TN_rate is zero and none of the minority instances is correctly predicted.

Table 4: *IRWMean and B-mean Calculation for Phishing Detection*

Imbalance Ratio (IR)	*1/IR*	*IRWMean*	*Acc*	*B-mean*
7.33	*0.136*	*0.016*	*0.91*	*0.46*

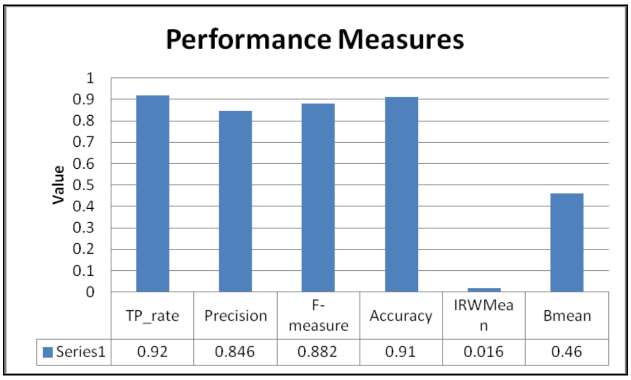

Figure 3: *Comparison of Performance Measures*

IRWMean is 0.016 which is close to the poor accuracy of minority class. IRWMean is an important parameter for performance evaluation when the misclassification cost of minority class is huge as compared to the majority class. The IRWMean gives true performance evaluation when there is a huge difference between the misclassification cost of minority and majority class. The B-mean gives more balanced performance evaluation based on the predictive accuracy of minority and majority classes and their misclassification cost. Considering the unbalanced nature and importance of both the classes of phishing datasets, the B-mean is more appropriate measure for performance evaluation.

Similar pattern of class unbalance is observed in medical datasets. In the dataset of cancer patients, if the classifier categorizes all patients as free from cancer, it automatically claims more than 90% of accuracy because in the cancer dataset cancer-free patients are in majority and cover more than 90% of the dataset. However, such classifiers cost life loss due to the misleading performance measures. In this type of applications of unbalanced data, IRWMean can give a more valid performance evaluation. In the medical domain, reliable, accurate and unbiased performance evaluation is important as misclassification of minority samples cost a loss of life.

6. Conclusion

Many researchers proposed works for the phishing detection. Most of the study claims secure and safe environment with more than 90% accuracy. However, in the case of class imbalance, high accuracy is misleading. The actual accuracy of minority class detection gets outnumbered by the high accuracy of the majority class that makes the security system unreliable. The accurate detection of the majority class, which covers large portion of the target dataset, influence the accuracy measure. In the case of the unbalanced dataset, where the misclassification cost of minority class is huge, the majority-biased misleading accuracy result into increased misclassification cost. Our experiment suggests that the performance measures should be carefully selected when the misclassification cost is uneven in unbalanced data classification. More than 90% of accuracy donot entitle any system to be reliable and secure. Comprehensive and generalized performance evaluation with unbiased measures is necessary for the reliable cyber security system. The B-mean and IRWMean performance measure consider the misclassification cost and gives un-biased and reliable performance evaluation. In the future, we will experiment with a medical dataset that generally observe large variations in the misclassification cost of minority and majority classes. In the medical domain reliability of machine learning based system is very important to save human lives.

References

[1] Jain AK, Gupta BB. PHISH-SAFE: URL features-based phishing detection system using machine learning. Adv Intell Syst Comput 2018;729:467–74. https://doi.org/10.1007/978-981-10-8536-9_44.

[2] Ankit Kumar Jain BBG. Towards detection of phishing websites on client-side using machine learning based approach. Springer Sci Bus Media 2017.

[3] Aaron G. Phishing Activity Trends Report 2nd Quarter. Anti-Phishing Work Gr 2019:1–12.

[4] Arachchilage NAG, Love S, Beznosov K. Phishing threat avoidance behaviour: An empirical investigation. Comput Human Behav 2016;60:185–97. https://doi.org/10.1016/j.chb.2016.02.065.

[5] Sheng S, Wardman B, Warner G, Cranor LF, Hong J, Zhang C. An empirical analysis of phishing blacklists. 6th Conf. Email Anti-Spam, CEAS 2009, 2009.

[6] Jain AK, Gupta BB. Phishing detection: Analysis of visual similarity based approaches. Secur Commun Networks 2017;2017. https://doi.org/10.1155/2017/5421046.

[7] Barot PA, Jethva HB. ImbTree: Minority Class Sensitive Weighted Decision Tree for Classification of Unbalanced Data. Int J Intell Syst Appl Eng 2021;9:152–8. https://doi.org/10.18201/IJISAE.2021473633.

[8] Barot P, Jethva H. MiNB: Minority Sensitive Naïve Bayesian Algorithm for Multi-Class Classification of Unbalanced Data. Int Arab J Inf Technol 2022;19:609–16. https://doi.org/10.34028/iajit/19/4/5.

[9] Chmielowski L, Konstantynov P, Luczak R, Kucharzak M, Burduk R. a Novel Method for Software Bug Report Assignment. Reliab Theory Appl 2023;18:579–88. https://doi.org/10.24412/1932-2321-2023-273-579-588.

[10] Sheng S, Magnien B, Kumaraguru P, Acquisti A, Cranor LF, Hong J, et al. Anti-Phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. ACM Int. Conf. Proceeding Ser., vol. 229, 2007, p. 88–99. https://doi.org/10.1145/1280680.1280692.

[11] Jain AK, Gupta BB. A novel approach to protect against phishing attacks at client side using auto-updated white-list. Eurasip J Inf Secur 2016;2016:1–11. https://doi.org/10.1186/s13635-016-0034-3.

[12] Nguyen LAT, To BL, Nguyen HK, Nguyen MH. Detecting phishing web sites: A heuristic URL-based approach. Int. Conf. Adv. Technol. Commun., 2013, p. 597–602. https://doi.org/10.1109/ATC.2013.6698185.

[13] Qabajeh I, Thabtah F, Chiclana F. A recent review of conventional vs. automated cybersecurity anti-phishing techniques. Comput Sci Rev 2018;29:44–55. https://doi.org/10.1016/j.cosrev.2018.05.003.

[14] Khonji M, Iraqi Y JA. Phishing Detection: A Literature Survey. IEEE Commun Surv Tutor 2013:2091–121.

[15] Abbas AR, Singh S, Kau M. Detection of phishing websites using machine learning. Lect Notes Networks Syst 2020;89:1307–14. https://doi.org/10.1007/978-981-15-0146-3_128.

[16] Subasi A, Molah E, Almkallawi F, Chaudhery TJ. Intelligent phishing website detection using random forest classifier. 2017 Int Conf Electr Comput Technol Appl ICECTA 2017 2017;2018-Janua:1–5. https://doi.org/10.1109/ICECTA.2017.8252051.

[17] Bayu Adhi Tama K-HR. A Comparative Study of Phishing Websites Classification Based on Classifier Ensembles. J Korea Multimed Soc 2018;21.

[18] Alejandro Correa Bahnsen ECB. Classifying Phishing URLs Using Recurrent Neural Networks. IEEE 2017.

[19] Chunlin Liu, Bo Lang, L Wang YZ. Finding effective classifier for malicious URL detection. ACM 2018.

[20] Pratik A Barot HBJ. Empirical Study of Evaluation Metric to Proposed New Balanced Metric (B-Mean) for Unbiased Performance Evaluation of Imbalanced Data Classification. IJRAR 2018.

[21] Gautam S, Rani K, Joshi B. Detecting phishing websites using rule-based classification

algorithm: a comparison. Lect Notes Networks Syst 2018;9:21–33. https://doi.org/10.1007/978-981-10-3932-4_3.

[22] Desai A, Jatakia J, Naik R, Raul N. Malicious web content detection using machine leaning. RTEICT 2017 - 2nd IEEE Int Conf Recent Trends Electron Inf Commun Technol Proc 2017;2018-Janua:1432–6. https://doi.org/10.1109/RTEICT.2017.8256834.

[23] Ahmad Abunadi, Anazida Zainal OA. Feature Extraction Process: A Phishing Detection Approach. IEEE 2013.

[24] M. Amaad Ul Haq Tahir, Sohail Asghar, Ayesha Zafar SG. A Hybrid Model to Detect Phishing-Sites using Supervised Learning Algorithms. Int. Conf. Comput. Sci. Comput. Intell. IEEE, 2016.

[25] E. B, K. T. Phishing URL Detection: A Machine Learning and Web Mining-based Approach. Int J Comput Appl 2015;123:46–50. https://doi.org/10.5120/ijca2015905665.

[26] Barot PA, Jethva HB. Statistical Study to Prove Importance of Causal Relationship Extraction in Rare Class Classi fi cation 2018;1. https://doi.org/10.1007/978-3-319-63673-3.

[27] BAROT PA, JETHVA HB. Mgini-improved decision tree using minority class sensitive splitting criterion for imbalanced data of covid-19. J Inf Sci Eng 2021;37:1097–108. https://doi.org/10.6688/JISE.202109_37(5).0008.

[28] Phishing dataset PhishTank n.d. https://www.phishtank.com/developer_info.php.

[29] Legitimate Url Dataset DMOZ n.d. https://www.dmoz.org/.