iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61274

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Phishing Website Detection

Mohd Shariyab[1], Praveen Singh[2], Dr. Sadhana Rana[3]
Computer Science and Engineering, SRMCEM, Lucknow, India

Abstract: ***Phishing is an online threat where an attacker impersonates an authentic and trustworthy organization to obtain sensitive information from a victim. One example of such is trolling, which has long been considered a problem. However, recent advances in phishing detection, such as machine learning-based methods, have assisted in combatting these attacks. Therefore, this paper develops and compares four models for investigating the efficiency of using machine learning to detect phishing domains. It also compares the most accurate model of the four with existing solutions in the literature. The work carried out in this study is an update in the previous systematic literature surveys with more focus on the latest trends in phishing detection techniques. This study enhances readers' understanding of different types of phishing website detection techniques, the data sets used, and the comparative performance of algorithms used. Our findings show that the model based on the K means clustering is the most accurate of the other four techniques and outperforms other solutions in the literature.***
Keywords: phishing detection, machine learning, phishing domains, artificial neural networks, support vector machine, decision tree, random forest.

I. INTRODUCTION

The rapid evolution of the technology has brought unpredictable convenience to our lives. But it has also given rise to a significant threat-phishing attacks. Social engineering attacks are common security threat which are used to reveal the private and confidential information by simply tricking the user without being detected. Phishing attacks are basically fraudulent emails, text messages, phone calls, websites that are designed to trick the user for downloading the malware and make user to share the sensitive information or they make user to share the personal data. Personal data can be anyone's bank account details, card numbers, any social media id or the login credentials. Phishing is the most common type of the social engineering attack. The practice of deceiving, pressuring or manipulating people into sending information or assets to the wrong people. Social engineering attacks rely on human error and pressuring tactics for the success. The attacker typically masquerades as a person or organization the victim trusts—e.g., a coworker, a boss, a company the victim or victim's employer does business with—and creates a sense of urgency that drives the victim to act rashly. Hackers and fraudsters use these tactics because it's easier and less expensive to trick people than it is to hack into a computer or network. Typically, phishing attack exploits the social engineering to lure the victim through sending a spoofed link by redirecting the victim to a fake web page. The spoofed link is placed on the popular web pages or sent via email to the victim. The fake webpage is created similar to the legitimate webpage. Thus, rather than directing the victim request to the real web server, it will be directed to the attacker server. The current solutions of antivirus, firewall and designated software do not fully prevent the web spoofing attack.

The implementation of Secure Socket Layer (SSL) and digital certificate (CA) also does not protect the web user against such attack. In web spoofing attack, the attacker diverts the request to fake web server. In fact, a certain type of SSL and CA can be forged while everything appears to be legitimate. According to, secure browsing connection does virtually nothing to protect the users especially from the attackers that have knowledge on how the "secure" connections actually work. This paper develops an anti-web spoofing solution based on inspecting the URLs of fake web pages. This solution developed series of steps to check characteristics of websites Uniform Resources Locators (URLs).

Our Phishing detection website project is a proactive response to the escalating cyber threats that exploits human vulnerability. The website is meticulously designed to combat phishing attempts by employing advanced algorithms, machine learning, and real-time data analysis. By leveraging these technologies, our platform will empower users to identify and thwart phishing attacks effectively, thereby safeguarding their sensitive information from failing into the wrong hands.

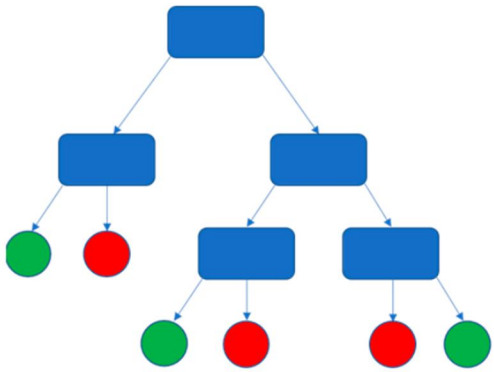
II. BACKGROUND

Some machine learning algorithms that are currently being used and have proven efficient in phishing domain detection, some of these are:

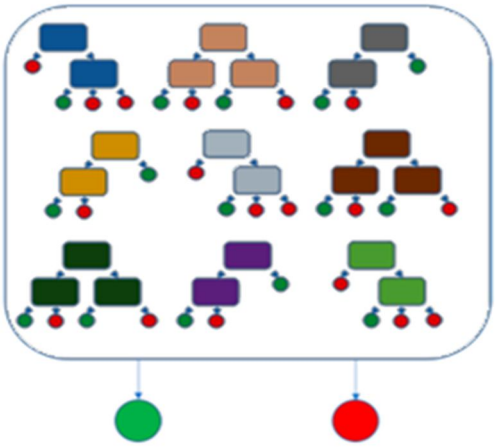
A. *Random Forest*

Random forest is a collection of supervised learning algorithms for classification and regression used in predictive modeling and machine learning [1]. Random forest has attracted attention due to its fast distribution and high accuracy. It aggregates the results and predictions of various decision trees to select the best results: class type (most common value in the decision tree) or average predictions. Random forest divides the data set into two parts: training and testing. It then randomly selects many examples from the training. Then, for each example, the researchers used a decision tree that divided each option into two children using the optimal distribution. After that, users must repeat the last step to vote for each prediction and choose the prediction with the most votes as the final result. The main hyperparameters in random forest are used to increase the predictive power of the model or make the model faster [2]. In this case more trees can improve performance and make predictions more stable, but can also increase processing time. Using the maximum number of pages in addition to the minimum number of pages can improve the performance of the algorithm. Once the training step is completed, the model can be applied to test data. This method allows results to be predicted and then compared to expected results [3]. Figure 2 shows how each tree is responsible for producing different products when given an independent random sample.

The random forest is used for its error generalization technique, and the random forest's accuracy improves as the forest grows in size. After randomly picking the features for the error rate, the accuracy is entirely dependent on the correlation between the trees. The random forest's characteristics might be created by tracking the error and correlation between nodes. As a consequence, the relevance of a variable can be measured.



Decision Tree



Random Forest

Figure 2. A comparison of DT and RF [4].

B. Support Vector Machine

SVM is a supervised learning method based on pattern recognition and regression study. Scientific research can identify the key factors needed to successfully learn specific, simple algorithms; Most applications in the world need to use complex tools and algorithms (such as neural networks); This is also very important in theory. It is difficult to define. SVM is the intersection of learning theory and practice. The models they create are both complex (for example, they feature a large class of neural networks) and yet simple enough to be analyzed mathematically. This is because SVM is a linear algorithm in high-dimensional space [5]. As shown in Figure 3 SVM predicts labels by creating a decision boundary (like a general plane) with at least one label between two groups. Data points and support vectors are controlled by hyperplanes. Uses the distance between data points to classify each group independently. Previous research has demonstrated that the hyperplane with the greatest margin of separation between the two classes offers the highest generalization performance [6].The best hyperplane is found by solving a convex optimization problem involving the minimization of a quadratic function under linear inequality constraints. The answer maybe expressed in terms of support vectors, which are a subset of the training instances. Support vectors include all the information required to solve a classification issue since the result will remain the same even if all other vectors are removed.

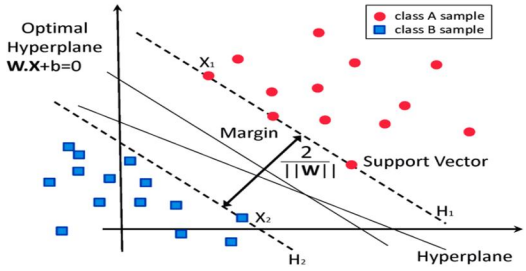


Figure 3. Support vector machine [2].

C. Gradient Boosting

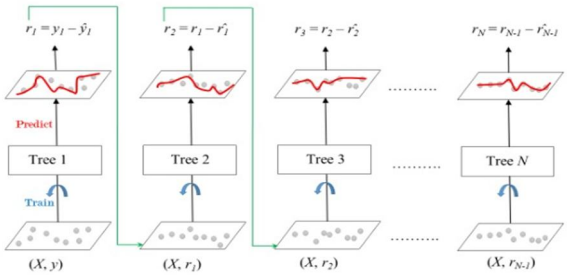
Gradient Boosting algorithms have emerged as a focal point in machine learning research owing to their exceptional performance across a wide range of predictive tasks. In research papers, Gradient Boosting is frequently scrutinized for its ability to enhance predictive accuracy, particularly when confronted with extensive and intricate datasets. Scholars often delve into the algorithm's nuances, proposing innovative enhancements such as novel loss functions, regularization methods, or optimization strategies to augment performance or tackle specific challenges like overfitting.

Moreover, the applicability of Gradient Boosting across diverse domains such as finance, healthcare, natural language processing, and computer vision is a common subject of investigation, with researchers examining its comparative efficacy against other machine learning techniques and tailoring its implementation to accommodate specific data characteristics or tasks.

As scalability can be a concern due to the sequential nature of Gradient Boosting, research papers frequently explore methods to improve efficiency, including parallelization, distributed computing, or hardware acceleration. Additionally, efforts to enhance the interpretability of Gradient Boosting models are prevalent, with researchers devising techniques such as feature importance analysis, partial dependence plots, and model visualization to elucidate the inner workings of these complex algorithms.

Through benchmarking and comparative studies, researchers aim to elucidate the strengths and weaknesses of Gradient Boosting, thus contributing to the advancement of machine learning methodologies and applications.

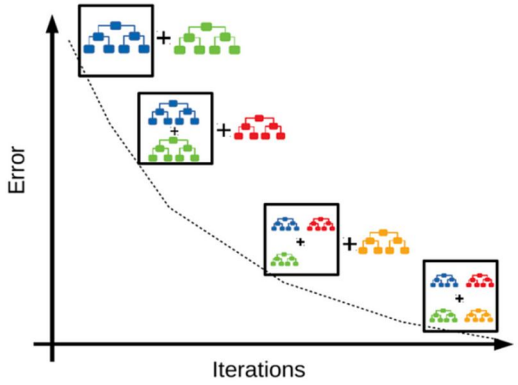
There is a technique called the Gradient Boosted Trees whose base learner is CART (Classification and Regression Trees). The below diagram explains how gradient-boosted trees are trained for regression problems.



The ensemble consists of M trees. Tree1 is trained using the feature matrix X and the labels y. The predictions labeled y1(hat) are used to determine the training set residual errors r1. Tree2 is then trained using the feature matrix X and the residual errors r1 of Tree1 as labels. The predicted results r1(hat) are then used to determine the residual r2. The process is repeated until all the M trees forming the ensemble are trained. There is an important parameter used in this technique known as Shrinkage. Shrinkage refers to the fact that the prediction of each tree in the ensemble is shrunk after it is multiplied by the learning rate (eta) which ranges between 0 to 1. There is a trade-off between eta and the number of estimators, decreasing learning rate needs to be compensated with increasing estimators in order to reach certain model performance. Since all trees are trained now, predictions can be made. Each tree predicts a label and the final prediction is given by the formula,

y(pred) =

y1 + (eta * r1) + (eta *r2)+…+(eta *rN)



D. Logistic Regression

Logistic regression uses a logistic function called a sigmoid function to map predictions and their probabilities. The sigmoid function refers to an S-shaped curve that converts any real value to a range between 0 and 1.

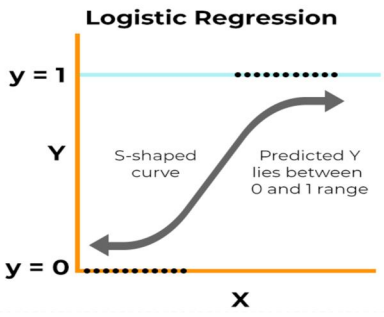
Moreover, if the output of the sigmoid function (estimated probability) is greater than a predefined threshold on the graph, the model predicts that the instance belongs to that class. If the estimated probability is less than the predefined threshold, the model predicts that the instance does not belong to the class.

The sigmoid function is referred to as an activation function for logistic regression and is defined as:

$$f(x) = \frac{1}{1 + e^{-x}}$$

where,

- e = base of natural logarithms
- value = numerical value one wishes to transform.



III. RELATED WORK

In general, users will ignore website URLs. This increases their chances of falling into phishing domains, which can be prevented by determining whether the URL is genuine. Unfortunately, modern methods for detecting phishing attacks have limited accuracy and detect only 20% of attempts. Machine learning techniques for phishing detection can produce better results, but they are time-consuming and not scalable even with small databases. Additionally, heuristic-based phishing detection has a false positive rate. Previous research on anti-phishing models has focused on strategies to change performance.

However, the use of reduced and integrated models can increase the accuracy of these models. Machine learning algorithms for phishing domain detection are popular and their use has become a simple classification problem. To build an ML detection model, the cell data must contain features related to phishing and legitimate websites in the cluster. Previous studies have shown that detection accuracy is high when using robust machine learning. Various selection strategies are used to reduce features. To train a machine learning model to predict phishing attacks and legitimate traffic, a dataset needs to be provided as input.

When features are reduced, dataset visualization becomes more efficient and easier to understand. The most important products of DT, C4.5, k-NN and SVM algorithms are; They have used many research projects and investigated phishing attacks with the most accurate and effective results. As empirical tests show, manually adjust parameters and training periods, and poor detection accuracy are prevalent problems.

Despite these benefits, researchers have noted the limits of their studies. Many pointed out that ensemble learning techniques have not been applied and that feature selection and reduction have not been performed. A range of strategies has been applied to combat phishing attacks. One paper [7] used different classifiers, such as naive Bayes and SVM. Similarly, the authors in [8] utilized random forest to differentiate phishing attacks from normal websites.

Table 1. Comparison table of the latest research focusing on machine learning phishing detection techniques

Model	Dataset	Algorithm	Accuracy
James et al. [37]	URLs	IBK, SVM, NB	89.75%
Subasi et al. [17]	website	ANN, KNN, RF, SVM, C4.5, RF	97.36%
Mao et al. [50]	Websites	SVM, RF, DT, AB	93%
Tyagi et al. [51]	URLs	DT, RF, GBM	98.40%
Chen and Chen [52]	websites	ELM, SVM, LR, C4.5, LC-ELM, KNN, XGB	99.2%
Joshi et al. [41]	Websites	RF	97.63%
Ubing et al. [42]	UCI	Ensemble bagging, boosting, stacking	95.4%
Sahingoz et al. [56]	Websites	SVM, DT, RF, KNN, KS, NB	97.98%
Abdelhamid et al. [53]	URLs	eDRI	93.5%
Patil et al. [40]	URLs	LR, DT, RF	96.58%
Jain and Gupta [54]	Websites	RF	99.57%
Jagadeesan et al. [57]	URLs	RF, SVM	95.11%
Niranjan et al. [58]	Websites	RC, kNN, IBK, LR, PART	97.3%
Chiew et al. [59]	URLs	RF, C4.5, PART, SVM, NB	96.17%
Pandey et al. [60]	Websites	SVM, RF	94%
Ali and Ahmed [61]	Websites	Genetic algorithm (GA) + DNN	89.50%
Aljofey et al. [62]	Websites	CNN	95.02%
Shie [63]	Websites	Convolutional auto encoder + DNN	89.00%
Maurya and Jain [64]	Websites	PSL 1 + PART	99.30%
Wang et al. [65]	Websites	RNN + CNN	95.79%
Lakshmi et al. [55]	UCI	DNN +Adam	96.00%
Li et al. [43]	URLs	GBDT, XGBoost and LightGBM	98.60%
Yang et al. [66]	Websites	Auto encoder + NIOSELM	94.60%
Anupam and Arpan [67]	Websites	Grey wolf optimizer + SVM	90.38%

IV. METHODOLOGY

Utilizing the Kaggle dataset, four phishing detection models were developed using K means clustering algorithms. The normalization feature was employed as a preprocessing strategy to improve the models' accuracy. The proposed models were able to detect different types of attacks from the UCI dataset. The following subsections discuss the dataset used and implemented algorithms; Sections 4.1 and 4.2, respectfully.

A. Dataset Used

The dataset is borrowed from Kaggle, https://www.kaggle.com/eswarchandt/phishing-website-detector. A collection of website URLs for 11000+ websites. Each sample has 30 website parameters and a class label identifying it as a phishing website or not (1 or -1). The overview of this dataset is, it has 11054 samples with 32 features.

B. Implemented Algorithm

To increase accuracy, this paper utilized the MinMax normalization feature as a preprocessing step in each proposed model. Normalization is a useful strategy for improving the accuracy of machine learning models, and it is required for some models to work properly. The MinMax normalization technique in the suggested model compresses the data to a domain of [0, 1], which improves the model training input quality (see Equations (1) and (2)).

$X\_std = (X - X.min) / (X.max - X.min)$ ……..(1)

$X\_scalar = X\_std \times (max - min) + min$ ……..(2)

To enhance the model performance and complexities, we used a data normalization strategy, as shown in Table 2. The algorithm selects significant aspects from the initial dataset by determining the prediction outcome, which is performed by filtering it through 30 features. The UCI dataset is split 80/20 into training and testing sets, respectively, by using c5-fold cross-validation, which presented the best performance in the latest research. The prediction model is then taught using machine learning, which employs various learning models. This is particularly useful for making predictions, as utilizing many models ensures that the results are not biased toward a single model. To account for this, we present the results of all the models combined and totaled to establish their maximum accuracies. If most of the models indicate that a domain is phishing, then the model's prediction accuracy confirms that the domain is a phishing attempt.

Classifier	Training	Testing
Gradient Boosting	Accuracy: 98.99% Precision: 98.66% Recall: 99.12% F1-measure: 98.99%	Accuracy: 97.78% Precision: 97.81% Recall: 98.21% F1-measure: 98.01%
SVM	Accuracy: 98.46% Precision: 98.24% Recall: 98.70% F1-measure: 98.47%	Accuracy: 97.06% Precision: 97.47% Recall: 97.24% F1-measure: 97.36%
Random Forest	Accuracy: 96.27% Precision: 95.61% Recall: 96.99% F1-measure: 96.30%	Accuracy: 95.52% Precision: 96.09% Recall: 95.86% F1-measure: 95.98%
Logistic Regression	Accuracy: 92.74% Precision: 91.93% Recall: 93.72% F1-measure: 92.81%	Accuracy: 92.89% Precision: 94.46% Recall: 92.70% F1-measure: 93.57%

V. MODEL'S FLOWCHART

Phishing is a concern to many individuals. However, existing methods, such as browser security indicators, cannot detect phishing websites. Due to the limits of current technology, users must evaluate whether a URL is phishing or not on their own. As a result, an automated technique for phishing website identification should be explored for increased cyber safety. This study shows how an implemented feature extraction approach and a prediction model based on a random forest classifier help increase the likelihood that a user will correctly identify a phishing website.

Each of the developed models, as shown in Figure 7, employs a feature selection technique to increase its accuracy. The data analysis heat map picks those that are most crucial in affecting the forecasted result by filtering the most interesting features out of the original dataset. As a result, irrelevant features have no effect on the model's efficiency or prediction.

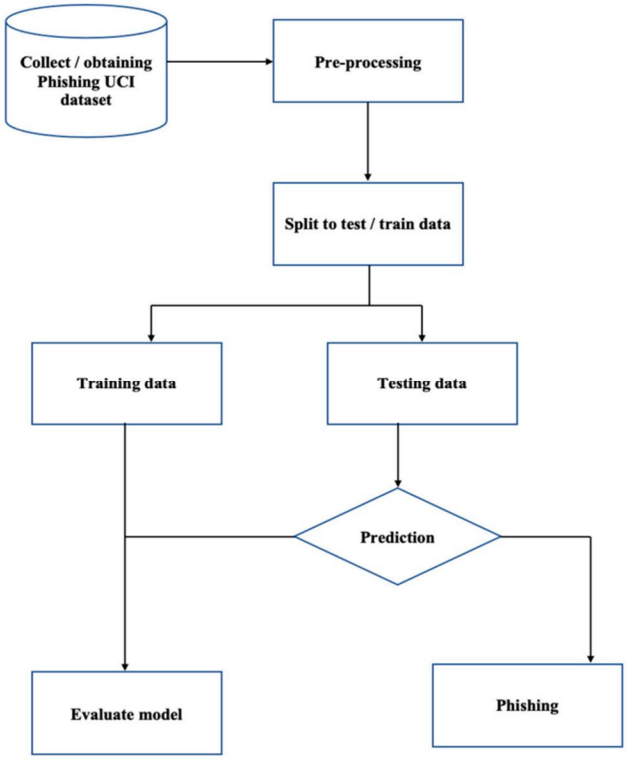


Figure 7. Model's flowchart.

VI. FUTURE WORKS

Phishing detection is a critical task in cybersecurity. Machine learning algorithms have been used to detect phishing websites with high accuracy. Future work in this area can focus on the following:

1) *Improving the Accuracy of Phishing Detection:* Researchers can explore new machine learning algorithms and techniques to improve the accuracy of phishing detection. For example, researchers can use deep learning algorithms to detect phishing websites based on their visual content.
2) *Detecting zero-day Phishing Attacks:* Zero-day phishing attacks are new and unknown attacks that have not been seen before. Researchers can develop machine learning algorithms that can detect zero-day phishing attacks by analyzing the behavior of users and the network.
3) *Detecting Phishing Attacks on mobile Devices:* With the increasing use of mobile devices, phishing attacks on mobile devices are becoming more common. Researchers can develop machine learning algorithms that can detect phishing attacks on mobile devices by analyzing the user's behavior and the characteristics of the mobile device.
4) *Developing real-time Phishing Detection Systems:* Real-time phishing detection systems can detect phishing attacks as they happen, allowing users to take immediate action to protect themselves. Researchers can develop machine learning algorithms that can detect phishing attacks in real-time by analyzing network traffic and user behavior.

VII. CONCLUSION

In this work, we investigated the practicality and the efficiency of using machine learning for phishing detection. We developed four machine learning models based on artificial neural networks (ANNs), support vector machines (SVMs), logistic regression, gradient boosting and random forest (RF) techniques. We then selected the most outperforming model of the fours and compared its performance with other solutions in the literature. The overall results show random forest (RF) model achieved the highest performance and outperforms other schemes in the literature.

The most important way to protect the user from phishing attack is the education awareness. Internet users must be aware of all security tips which are given by experts. Every user should also be trained not to blindly follow the links to websites where they have to enter their sensitive information. It is essential to check the URL before entering the website. In Future System can upgrade to automatic Detect the web page and the compatibility of the Application with the web browser. Additional work also can be done by adding some other characteristics to distinguishing the fake web pages from the legitimate web pages. PhishChecker application also can be upgraded into the web phone application in detecting phishing on the mobile platform.

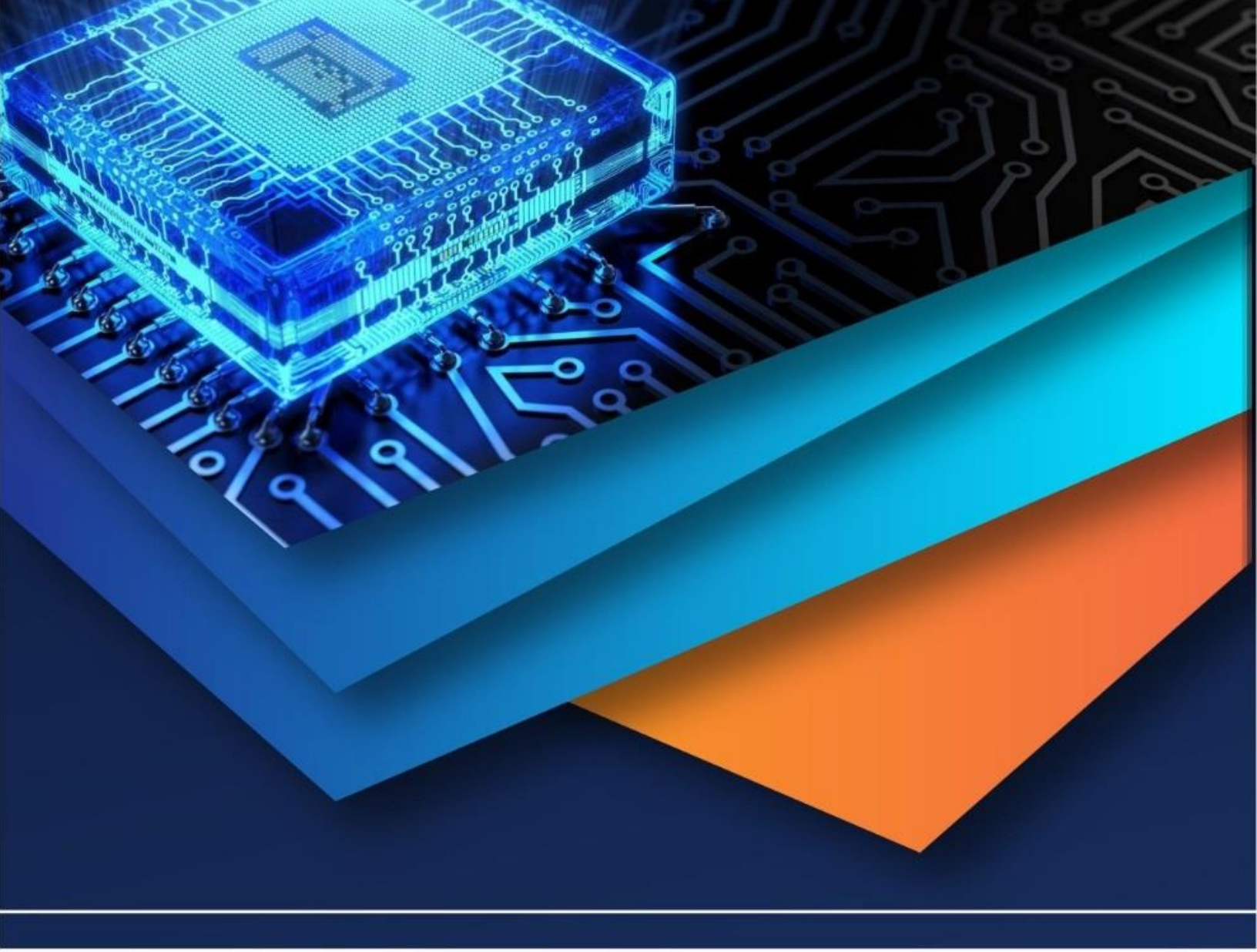
There are many features that can be improved in the work, for various other issues. The heuristics can be further developed to detect phishing attacks in the presence of embedded objects like flash. Identity extraction is an important operation and it was improved with the Optical Character Recognition (OCR) system to extract the text and images. More effective inferring rules for identifying a given suspicious web page, and strategies for discovering if it is a phishing target, should be designed in order to further improve the overall performance of this system.

Moreover, it is an open challenge to develop a robust malware detection method, retaining accuracy for future phishing emails. In addition, the dynamic and static features complement each other, and therefore both are considered important in achieving high accuracy.

REFERENCES

[1] Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
[2] Friedman, J.H. The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Springer Open: Berlin/Heidelberg, Germany, 2017.
[3] Brownlee, J. Train-Test Split for Evaluating Machine Learning Algorithms. Mach. Learn. Mastery 2020, 23. Available online: https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/ (accessed on 25 December 2021).
[4] Jeremybeauchamp English: A Visual Comparison between the Complexity of Decision Trees and Random Forests. 2020. Available online: https://commons.wikimedia.org/wiki/File:Decision_Tree_vs._Random_Forest.png (accessed on 27 December 2021).
[5] Sönmez, Y.; Tuncer, T.; Gökal, H.; Avcı, E. Phishing Web Sites Features Classification Based on Extreme Learning Machine. In Proceedings of the 2018 6th International Symposium on Digital Forensic and Security (ISDFS), IEEE, Antalya, Turkey, 22–25 March 2018; pp. 1–5.
[6] Cristianini, N.; Shawe-Taylor, J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods; Cambridge University Press: Cambridge, UK, 2000.
[7] James, J.; Sandhya, L.; Thomas, C. Detection of Phishing URLs Using Machine Learning Techniques. In Proceedings of the 2013 International Conference on Control Communication and Computing (ICCC), Thiruvananthapuram, India, 13–15 December 2013; Available online: https://ieeexplore.ieee.org/abstract/document/6731669 (accessed on 26 September 2021).
[8] Liew, S.W.; Sani NF, M.; Abdullah, M.T.; Yaakob, R.; Sharum, M.Y. An Effective Security Alert Mechanism for Real-Time Phishing Tweet Detection on Twitter—ScienceDirect. Comput. Secur. 2019, 83, 201–207. Available online: https://www.sciencedirect.com/science/article/pii/S0167404818309040 (accessed on 26 September 2021). [CrossRef]
[9] Hutchinson, S.; Zhang, Z.; Liu, Q. Detecting Phishing Websites with Random Forest. In Proceedings of the Machine Learning and Intelligent Communications, Hangzhou, China, 6–8 July 2018; Meng, L., Zhang, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 470–479.
[10] 40. Patil, V.; Thakkar, P.; Shah, C.; Bhat, T.; Godse, S.P. Detection and Prevention of Phishing Websites Using Machine Learning Approach. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 19–18 August 2018; pp. 1–5.
[11] Joshi, A.; Pattanshetti, P.T.R. Phishing Attack Detection Using Feature Selection Techniques; Social Science Research Network: Rochester, NY, USA, 2019.
[12] Ubing, A.; Kamilia, S.; Abdullah, A.; Zaman, N.; Supramaniam, M. Phishing Website Detection: An Improved Accuracy through Feature Selection and Ensemble Learning. Int. J. Adv. Comput. Sci. Appl. 2019, 10, 252–257. [CrossRef]
[13] Li, Y.; Yang, Z.; Chen, X.; Yuan, H.; Liu, W. A Stacking Model Using URL and HTML Features for Phishing Webpage Detection. Future Gener. Comput. Syst. 2019, 94, 27–39. [CrossRef]
[14] 44. Zamir, A.; Khan, H.U.; Iqbal, T.; Yousaf, N.; Aslam, F.; Anjum, A.; Hamdani, M. Phishing Web Site Detection Using Diverse Machine Learning Algorithms. Electron. Libr. 2020, 38, 65–80. [CrossRef]
[15] Alsariera, Y.A.; Adeyemo, V.E.; Balogun, A.O.; Alazzawi, A.K. AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites. IEEE Access 2020, 8, 142532–142542. [CrossRef]
[16] Ali, W.; Malebary, S. Particle Swarm Optimization-Based Feature Weighting for Improving Intelligent Phishing Website Detection. IEEE Access 2020, 8, 116766–116780. [CrossRef]
[17] Adebowale, M.A.; Lwin, K.T.; Sanchez, E.; Hossain, M.A. Intelligent Web-Phishing Detection and Protection Scheme Using Integrated Features of Images, Frames and Text—ScienceDirect. Expert Syst. Appl. 2019, 115, 300–313. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0957417418304925 (accessed on 26 September 2021). [CrossRef]
[18] El Aassal, A.; Baki, S.; Das, A.; Verma, R.M. An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs. IEEE Access 2020, 8, 22170–22192. Available online: https://ieeexplore.ieee.org/abstract/document/8970564 (accessed on 27 September 2021). [CrossRef]
[19] Subasi, A.; Kremic, E. Comparison of Adaboost with MultiBoosting for Phishing Website Detection—ScienceDirect. Procedia Comput. Sci. 2020, 168, 272–278. Available online: https://www.sciencedirect.com/science/article/pii/S1877050920303902 (accessed on 27 September 2021). [CrossRef]
[20] Mao, J.; Bian, J.; Tian, W.; Zhu, S.; Wei, T.; Li, A.; Liang, Z. Phishing Page Detection via Learning Classifiers from Page Layout Feature. EURASIP J. Wirel. Commun. Netw. 2019, 2019, 43. Available online: https://jwcn-eurasipjournals.springeropen.com/articles/10.1186/s13638-019-1361-0 (accessed on 27 September 2021). [CrossRef]

[21] A Novel Machine Learning Approach to Detect Phishing Websites. Available online: https://ieeexplore.ieee.org/abstract/document/8474040/ (accessed on 27 September 2021).
[22] Chen, Y.H.; Chen, J.L. AI@ntiPhish—Machine Learning Mechanisms for Cyber-Phishing Attack. IEICE Trans. Inf. Syst. 2019, 102, 878–887. Available online: https://www.jstage.jst.go.jp/article/transinf/E102.D/5/E102.D_2018NTI0001/_article/-char/ja/ (accessed on 27 September 2021). [CrossRef]
[23] Abdelhamid, N.; Thabtah, F.; Abdel-Jaber, H. Phishing Detection: A Recent Intelligent Machine Learning Comparison Based on Models Content and Features. In Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics, Beijing, China, 22–24 July 2017; Available online: https://ieeexplore.ieee.org/abstract/document/8004877 (accessed on 27 September 2021).
[24] Jain, A.K.; Gupta, B.B. Towards Detection of Phishing Websites on Client-Side Using Machine Learning Based Approach. Telecommun. Syst. 2018, 68, 687–700. Available online: https://link.springer.com/article/10.1007/s11235-017-0414-0 (accessed on 27 September 2021). [CrossRef]
[25] Lakshmi, L.; Reddy, M.P.; Santhaiah, C.; Reddy, U.J. Smart Phishing Detection in Web Pages Using Supervised Deep Learning Classification and Optimization Technique ADAM. Wirel. Pers. Commun. 2021, 118, 3549–3564. [CrossRef]
[26] Sahingoz, O.K.; Buber, E.; Demir, O.; Diri, B. Machine Learning Based Phishing Detection from URLs—ScienceDirect. Expert Syst. Appl. 2019, 117, 345–357. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0957417418306067 (accessed on 27 September 2021). [CrossRef]
[27] Jagadeesan, S. URL Phishing Analysis Using Random Forest. Int. J. Pure Appl. Math. 2018, 118, 4159–4163.
[28] Niranjan, A.; Haripriya, D.K.; Pooja, R.; Sarah, S.; Deepa Shenoy, P.; Venugopal, K.R. EKRV: Ensemble of KNN and Random Committee Using Voting for Efficient Classification of Phishing; Springer: Singapore, 2019; Available online: https://link.springer.com/chapter/10.1007/978-981-13-1708-8_37 (accessed on 27 September 2021).
[29] Chiew, K.L.; Tan, C.L.; Wong, K.; Yong, K.S.; Tiong, W.K. A New Hybrid Ensemble Feature Selection Framework for Machine Learning-Based Phishing Detection System—ScienceDirect. Inf. Sci. 2019, 484, 153–166. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0020025519300763 (accessed on 27 September 2021). [CrossRef]
[30] Pandey, A.; Gill, N.; Sai Prasad Nadendla, K.; Thaseen, I.S. Identification of Phishing Attack in Websites Using Random ForestSVM Hybrid Model. In Proceedings of the Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018), Vellore, India, 6–8 December 2018; Springer International Publishing: Midtown Manhattan, NY, USA, 2020. Available online: https://link.springer.com/chapter/0.1007/978-3-030-16660-1_12 (accessed on 27 September 2021).
[31] Ali, W.; Ahmed, A.A. Hybrid Intelligent Phishing Website Prediction Using Deep Neural Networks with Genetic Algorithm-Based Feature Selection and Weighting. IET Inf. Secur. 2019, 13, 659–669. [CrossRef]
[32] Aljofey, A.; Jiang, Q.; Qu, Q.; Huang, M.; Niyigena, J.P. An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL. Electronics 2020, 9, 1514. Available online: https://www.mdpi.com/2079-9292/9/9/1514 (accessed on 27 September 2021). [CrossRef]
[33] Shie, E.W.S. Critical Analysis of Current Research Aimed at Improving Detection of Phishing 78Attacks. Sel. Comput. Res. Pap. 2020, 45, 45–53.
[34] Maurya, S.; Jain, A. Deep Learning to Combat Phishing. J. Stat. Manag. Syst. 2020, 23, 945–957. [CrossRef]
[35] Mao, J.; Bian, J.; Tian, W.; Zhu, S.; Wei, T.; Li, A.; Liang, Z. Detecting Phishing Websites via Aggregation Analysis of Page Layouts—ScienceDirect. Procedia Comput. 2018, 129, 224–230. Available online: https://www.sciencedirect.com/science/article/pii/S187705091830276X (accessed on 27 September 2021). [CrossRef]
[36] Yang, L.; Zhang, J.; Wang, X.; Li, Z.; Li, Z.; He, Y. An Improved ELM-Based and Data Preprocessing Integrated Approach for Phishing Detection Considering Comprehensive Features—ScienceDirect. Expert Syst. Appl. 2021, 165, 113863. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0957417420306734 (accessed on 27 September 2021). [CrossRef]



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY