# The Comparative Analysis of Machine Learning Algorithms for Phishing Attack Detection

Almina Sehrish[1], Muhammad Raza[2*], Muhammad Faizan Khan[3], Sabih Hida[4], Rageshwari Haryani[5]

***Abstract:* Due to the internet's indispensable role in day-to-day operations, cybercrimes have increased dramatically, with phishing being a serious concern. Phishing attacks employ phony websites to obtain sensitive data and user passwords. Because hackers are always trying to change their strategies wisely, traditional preventative measures like software detection and user awareness frequently fall short. With their capacity for self-learning, machine learning-based solutions provide a more potent protection. Using a supervised learning framework, this paper offers a thorough review of machine learning techniques for phishing detection, making use of a dataset that has 87 attributes. Runtime, train accuracy, test accuracy, precision, recall, and F1-score are used to assess the efficiency of various algorithms, such as Random Forest, Decision Trees, SVM, Naive Bayes, K-Nearest Neighbors (KNN), XGBoost, Boosted Decision Tree, AdaBoost, Extra Trees, LightGBM, and CatBoost. With a 97.33% test accuracy as well as outstanding recall, precision, and F1-score, XGBoost stands out. Powerful performance is also demonstrated by Random Forest and LightGBM, demonstrating their effectiveness in identifying phishing attempts. This study feeds future research on optimization tactics and ensemble approaches to improve detection robustness and accuracy, and it gives cybersecurity professionals effective insights for better phishing detection.**

***Keywords*: Phishing Detection, Machine Learning Algorithms, Supervised Learning, URL Analysis, Feature Extraction, Cybersecurity.**

## INTRODUCTION

Phishing attacks remain and will continue being a threat within the field of cybersecurity as a result of the following factors. These evil plans, which exploits the intentions of the users in the intention of receiving private information, is a serious threat to the world that is operated through electronic and digital platforms [1]. These activities are unwelcome and maligned, hence why strong detection methods are necessary to curb these actions as phishing strategies become sophisticated.

Machine learning algorithms have become an effective tool in combating phishing because the approach has the ability to learn from the data to find patterns and change its strategies as it adapts to new threats. Analyzing the up-to-date research by the [2], it can be stated that 78% of phishing websites have employed the SSL security protection while in the past this element was primarily associated with legally acting websites. This means that hackers have changed their operation style and made it more professional and polished. Phishing activity is not reduced in the second quarter as expected, but there is some rise in the last quarter of the year. According to (Figure 1) the APWG in the fourth quarter of the year 2023, an incidence of 1,077,501 in the number of phishing attacks was reported. 2023 marked as being the most Phishing prone year with an average of, five million phishing cases according to the APWG.

Different types of phishing are used as explained welled by [3].Phishing is a form of fraud, used through e-mail, its main aim is to trap people into providing personal details, such as credit card details and/or a password. Spear phishing is a form of phishing that is extremely specific and has been developed to give the appearance of a genuine communication to the person or company it is being sent to. Phishing attack is a form of whaling attack in which a hacker aims to gather crucial information from a recognized personality such as government official or a company official. Angler phishing is different scam style in which the hackers impersonate victims in such a way that they are compelled to reveal their personal details via social sites or online forums. There is a form of Malware that engages in the distribution of malware in web adverts; this is usually garnered by consumers clicking on links that lead to the infection of their gadgets. A watering hole attack makes it very easy for malware to invade the targeted users' devices by exploiting websites that they often frequent.

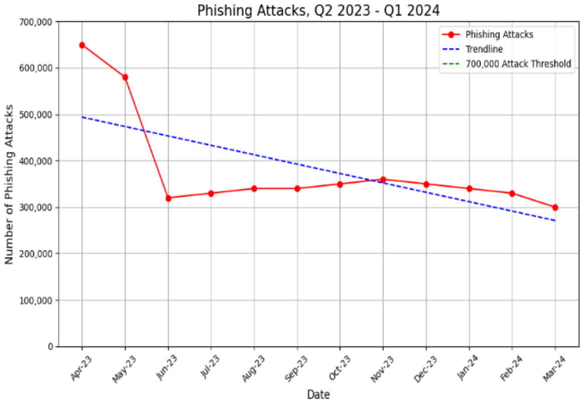

**Figure 1: Phishing Attacks [2]**

[1-4-5] SZABIST University
[2]SZABIST University, Gharo Campus
[3]IQRA University
Country: Pakistan
Email: *razacom_2000@yahoo.com

That is, it can be said that such Fourier machine learning tools are needed to safeguard consumers within the social web. This is a measure that maintains the privacy of the user's information and leads to safer web surfing by making the user aware of risks that are likely to be encountered while browsing and directing the user to the safer URLs. With these improvements, these users can bring out the experiences of utilizing the internet more easily and safely and reduce their exposures towards lying practices in internet marketing.

## PROBLEM STATEMENT

Phishing attacks using fake web addresses in mailing system increases data breach which leads assets and reputation losses.

One primary concern that the recent increase in web cloaking in phishing attacks in online systems has posed is the significantly increased threat of data compromise. The consequences outstretch farther than the primary victory of data, which includes asset loss and brand degradation of individuals and companies involved. The need to focus on this issue arises from the increasing concern with protecting data, potential adverse financial impacts, and lack of trust in organizations and businesses. Eliminating the rising incidents of phishing attacks, and the consequent misuse of fake web addresses is a crucial endeavor in abating data breaches and mitigating the multifaceted issues within physical and financial, human, informational, and reputational assets.

## SOLUTION STATEMENT

Supervised Machine learning Algorithms of cybersecurity to detect and block phishing attacks with fake web addresses in emails, minimizing data breaches and safeguarding assets and reputation.

To address the escalating risk of data breaches stemming from phishing attacks and the misuse of fake web addresses, this thesis proposes a comprehensive comparative analysis of machine learning algorithms for phishing detection. By leveraging advanced machine learning techniques, such as supervised learning, the study aims to identify and evaluate the most effective algorithms for detecting and thwarting phishing attempts in real-time. The goal is to equip cybersecurity practitioners and decision-makers with actionable insights to enhance their defenses against phishing attacks and safeguard sensitive information. Through empirical evidence and practical guidance, this research seeks to contribute to the development of initiative-taking measures to prevent data breaches and protect the integrity of digital ecosystems.

## LITERATURE REVIEW

The methods for knowing machine learning and deep learning for the purpose of protecting people against phishing are listed in [3] with their established potential on classification and regression abilities. Specific techniques that have been used include Random Forest, Decision Tree, Support Vector Machine, K-nearest neighbor, Logistic Regression, and AdaBoost. This emphasis accentuates how basic measures of performance such as accuracy, F-score, recall, and precision are when defining the efficiency of the system. It discusses the application of new techniques of the machine learning like Random Forests and neural networks for accurate formation of phishing, the classification of the phishing detection models concerning principal Feature selection and the deep learning techniques. It also focuses on the shift from capturing a deeper level of learning to the traditional learning technique showing the effectiveness of the deeper learning in profiling the latent linkages and making separate decisions.

In the paper titled, “Phish Haven—An Advanced Real-Time AI Phishing URLs Detection System,” [4] describe Phish Haven as a tool that has an artificial intelligence system able to accurately detect phishing URLs. As a URL Hit tool with Features Extractor and Modeless along with component Decision Maker, Phish Haven is helpful for extracting and filtering URLs. Phish Haven enhances initial letter and number counting lexicaling approaches through machine learning methods like logistic regression and neural networks. The numerous tests and scenarios reinforce Phish Haven as a very efficient tool in the current market, with the ability to work in near real-time, and therefore, its reassessing as one of the best cybersecurity solutions available.

This paper presents a novel adaptive technique for preventing distributed denial of service (DDoS) attacks in distributed environments. The technique leverages machine learning algorithms to detect and mitigate DDoS attacks by analyzing network traffic patterns and identifying malicious activities. The study highlights the effectiveness of the adaptive approach in reducing the impact of DDoS attacks and improving the overall security of distributed systems. The findings underscore the potential of machine learning in enhancing cybersecurity measures and protecting against evolving cyber threats [42].

[5] Pro side a comprehensive systematic analysis of the different research strategies employed for mitigating phishing utilizing Deep Learning algorithms. Drawing from the existing knowledge in the field, they found a knowledge gap particularly, Deep Learning strategies for identifying phishing attacks. The review highlighted 19 papers from 2014 to 2019 where authors discussed solely the works on the subject of phishing and Deep Learning interface.

This paper introduces a SYN Flood Attack Detection and Prevention Technique (SFaDMT) designed for distributed environments. The proposed method utilizes machine learning algorithms to analyze network traffic and identify potential SYN flood attacks. The study showcases the effectiveness of SFaDMT in detecting and preventing SYN flood attacks, thereby safeguarding distributed networks from malicious activities. The findings highlight the critical role of machine learning in developing robust cybersecurity solutions and protecting distributed systems from evolving cyber threats [43].

The review [6] indicated that researchers most frequently utilized the Random Forest Classifier, with the Support Vector Machine (SVM) and Decision Tree algorithms also being commonly employed. While the study [7] concludes that among the four models evaluated, the random forest (RF) model demonstrated superior performance, surpassing other approaches documented in the literature.

Thus, [8] provided an extensive analysis of the phishing and anti-phishing schemes, giving an insight into the most often utilised scams, including phone phishing, email impersonation, spear phishing, and email deception. Analyzing their systematic literature review (SLR), they found that the machine learning solution had the highest accuracy towards the phishing detection process.

This research focuses on detecting SYN flood attacks in Mobile Ad Hoc Networks (MANET) using a Bayesian Estimator-based approach (DsFaBe). The study evaluates the performance of the proposed method in identifying SYN flood attacks by analyzing network traffic data. The results demonstrate the efficacy of the Bayesian Estimator in accurately detecting and mitigating SYN flood attacks, thereby enhancing the security of MANETs. The paper emphasizes the importance of employing machine learning algorithms to address the unique challenges posed by the dynamic and decentralized nature of MANETs [44].

Nevertheless, one might state that their conclusion was derived from a small number of samples amounting to twenty studies, therefore, may well not applicable to all.
The review [9] indicated that researchers most frequently utilized the Random Forest Classifier, with the Support Vector Machine (SVM) and Decision Tree algorithms also being commonly employed and achieved CNN with accuracy of 99.98%.

This research [10] presents a novel approach that combines Random Forest (RF) and Convolutional Neural Networks (CNN) for phishing website detection. Therefore, in the given dataset, it is possible to find 11,430 URLs and 87 characteristics. The second objective aimed at utilizing it as the benchmark in machine learning-based phishing detection systems. The attributes are divided into three categories: 56 from the structural and syntactical patterns of the URLs respectively, 24 from content similarity in related web pages and 7 from the external databases by performing queries [13]. As pointed out in this paper, the sample is equally split, with each site, phishing and legitimate, occupying 50 % of the sample. In addition to the software, Python scripts for feature extraction are included for evaluating the work or replication purposes. These datasets were developed in May 2020 and have been updated to the latest versions requiring access to site content or outside services, it predicts the authenticity of a URL.

The study [11] addresses the benefits and drawbacks of several approaches, such as Random Forest, Adaboost, ISHO in conjunction with SVM, Random Forest, meta-learning algorithms, Convolutional Neural Networks (CNN), and others. These approaches have proven to be quite accurate in identifying phishing URLs, with accuracy rates ranging from 89% to 99.57%. While the work [12] provides a rules-based approach for phishing detection, using three machine learning models trained on a dataset of fourteen (14) features. The machine methods for learning include k-Nearest Neighbor (KNN), Random Forest, and Support Vector Machine. The Random Forest model outperformed the other two algorithms analyzed.

The application of machine learning to cybersecurity is covered in the paper [42], with an emphasis on DDoS attack detection and mitigation through the use of techniques like anomaly detection and adaptive learning. It draws attention to the advantages of federated learning, which enables decentralized detection procedures, reducing computational demands while maintaining efficient threat monitoring. The study also emphasizes how crucial it is to integrate these cutting-edge machine learning methods into existing cybersecurity frameworks, especially in complex contexts like cyber-physical systems, in order to increase overall security efficacy.

SYN flood attacks challenge ISPs by evading traditional detection. A study explains [43] CNN (94.2% accuracy), RNN (91.5%), and LSTM (96.0%) as efficient models, with LSTM offering the best detection but record computational cost, highlighting a trade-off between correctness and effectiveness for practical ISP utilization.

## METHODOLOGY

The technique for the "Comparative Analysis of Machine Learning Algorithms for Phishing Attack Detection" thesis begins with the installation of required Python libraries, such as CatBoost, LightGBM, and XGBoost, for developing machine learning algorithms. The relevant libraries are then loaded into the Python environment to help with activities like data manipulation, model training, evaluation, and visualization. Following that, the dataset is loaded into a Pandas DataFrame, a data structure offered by the Pandas package for efficient data processing and manipulation in Python.

The data is then cleaned taking into consideration missing values and measures are promptly put in place to ensure the integrity of data to be used throughout the study. Other necessary information checks include checking the balance of the classes within the dataset since the same is vital to prevent cases where some classes are dominant during training, rendering other classes irrelevant.

It represents a balanced state of attacks by using Python’s matplotlib and Seaborn, with a clear evaluation of the class distribution in the dataset. The next step involves the feature and target variable split and the target variable which is categorical is encoded into numeric format using python’s scikit-learn module since it is in the format suitable for machine learning algorithms. The input features are normalized where the data is scaled to have zero mean and unit variance for all input features, to make the scale of each input feature more uniform and to avoid data suffering from the curse of dimensionality where the features with larger magnitudes swamp the smaller ones.

Moreover, this dataset is split into training and testing datasets using the python scikit learn module to train and check the validity of the model to select the best model that can have high accuracy on unseen chunks of data. A set of classifiers is trained for every data set using Python ML libraries like scikit-learn, XGBoost, LightGBM, and CatBoost along with tuning parameters for a regularization to obtain consistent and generalized models. Individual classifiers are then built and trained with the training dataset and their performance is analysed using the metrics which include, accuracy, precision, recall and F1-score for the purpose of phishing detection using Python language scikit-learn module.

Thus, the final results can be represented in the confusion matrix for each classifier with the help of Python's matplotlib package. Machine learning techniques for Phishing attacks detection in Python as shown in Figure 2.

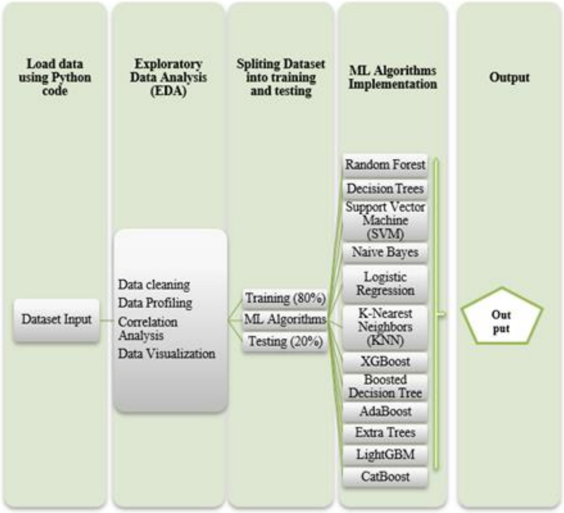

**Figure 2: Methodology of Phishing Detection Using Machine Learning Algorithm**

### *Evaluating Performance in Phishing Detection*

However, before short listing the various proposed phishing detection algorithms or models, a mathematical groundwork that can be used to measure its usefulness and measurability has to be established. As such, there is a use of such measures as true positives (TP), false positives (FP), true negatives (TN), and the invariably critical false negatives (FN).

### *Confusion Matrix*

This is sometimes referred to as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [10].

*True Positive (TP):* The values expected look like the actual values they reflect. Although, the actual result was positive but the value of the model in this case was also in positive sense.

*True Negative (TN):* According to anticipation, the value reflects the true value of the particular product that is to be manufactured. The number of links in turn was negative and was also expected by the model.

*False Positive (FP):* An α error. The actual value was over/underestimated or in other ways estimated falsely. It was even negative as opposed to the fact that the model learns gave a positive value.

*False negative (FN):* It is a type 2 mistake Moreover, the variables of such a concept are much more difficult to identify and define than in a type 1 mistake since it does not refer to fixed behavioral patterns or ways of acting. The predicted value is not the true value as is a projection of what the value should be in the future. While the value of the model was negative, in contrast, the true value was positive.

| | **Phishing-1** | **Legitimate-0** |
|---|---|---|
| **Phishing-1** | True Positive (TP) | False Positive (FN)<br>Type 1 Error |
| **Legitimate-0** | False Negative (FP)<br>Type 2 Error | True Negative (TN) |

**Figure 1: Confusion Matrix**

Some of the principles used in evaluating the performance of machine learning-based systems for phishing detection include recall, precision, F-score, and accuracy. Metrics like accuracy, precision, or recall can be obtained from a confusion matrix which includes TP, TN, FP, and FN.

$$\textit{Equation 1:}\quad Accuracy = \frac{\boldsymbol{T\,P + T\,N}}{\boldsymbol{T\,P + T\,N + FP + FN}}$$

Recall below (Equation 2) specifically assesses how accurately a system identifies positive samples. A higher recall means that even more, the samples that are regarded positive by the expert is accurately picked from the set. It is being estimated by using the formula of accuracy, where the number of samples that are correctly identified as positives divided by the number of actual positives.

$$\textit{Equation 2:}\quad Recall = \frac{\boldsymbol{TP}}{(\boldsymbol{TP+FN})}$$

Precision below (Equation 3) measures the accuracy of a system's positive predictions by calculating the ratio of correctly identified positive instances (true positives) to the total number of instances classified as positive, which includes both true positives and false positives.

$$\textit{Equation 3:}\quad Precision = \frac{\boldsymbol{TP}}{(\boldsymbol{TP+FP})}$$

The F1 test (Equation 4) is further referred to as the F-measure or the balanced F-score and is used as the highly sensitive criterion for evaluating the algorithms. It assesses the models by averaging the accuracy rate for a minor positive class with the specificity rate in a harmony mean.

$$\textit{Equation 4:}\quad F=2*\frac{\boldsymbol{Precision*Recall}}{\boldsymbol{Precision+Recall}}$$

Accuracy in above (Equation 1) can be defined as the way how a certain system works looking at the number of true positive and true negative classifications, therefore it offers a general outlook. In other words, when equal weight age is given to all the classes then this metric can be of immense importance. And it is calculated by using the formula: the total number of accurate predictions given by the system is divided by the total number of predictions issued by the system.

## RESULTS AND DISCUSSIONS

From the comparison analysis (Figure 4) it has been identified that advanced ensemble techniques XGBoost, LightGBM, CatBoost provide superior performance compared to other algorithms in the context of phishing detection having high accuracy and AUC at each iteration for different parameters. Some of the pre-existing methods like Random Forest, Gradient Boosting as well as SVM works well on the dataset in context, however models like Naive Bayes seem to have limitations here. These observations are particularly important for assessing the reliability and performance of the existing and potential future tools in the sphere of cybersecurity, namely in the detection of phishing attacks.

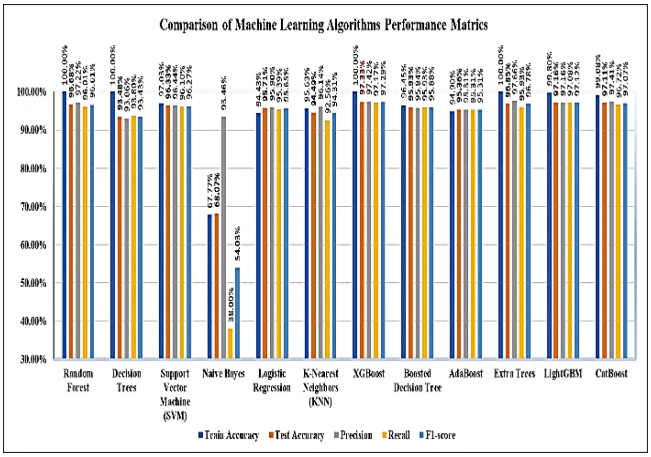

**Figure 4: Results of ML Algorithms**

### *Comparison of Algorithm Runtimes*

These findings highlight (Figure 5) the computational complexity for each algorithm, in terms of time required for computation. The runtimes of the algorithms are Naive Bayes, Logistic Regression and XGBoost are almost of the same, whereas the CatBOOST is observed to be taking longer time than all the listed algorithms. In the case of your application, given that you have certain requirements and constraints to be met, it would be advised that you choose an algorithm that provides the best balance of computational complexity and thoroughness of the algorithm.

### *Algorithm Runtimes ( Seconds)*

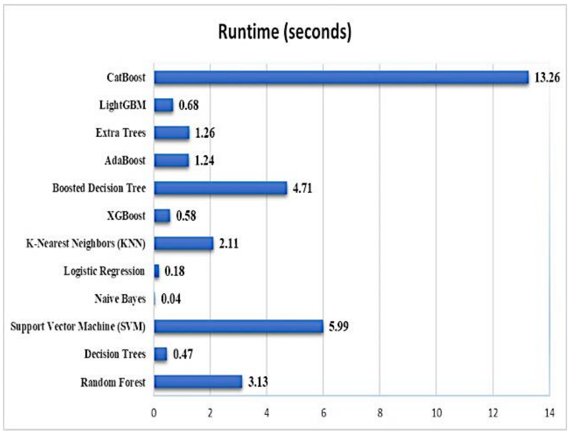

**Figure 2: Runtime of Algorithms**

In the comparative analysis (Table 1), the details of the study samples analyzed through various machine learning algorithms for identifying the incidence of phishing attacks are presented and evaluated in this section.

XGBoost performs well when evaluated because boosting technique is robust that enhances the ability to rectify mistakes while has a unique way of handling overfitting by means of regularization. The positivity of it is that it can optimize its performance and features to interact with other features making it best suited for large datasets. Naive Bayes have particularly poor results because it assumes that all the features are independent, which is rarely the case, leading to inferior results again. Additionally, its basic structure and ability to work well with continuous inputs and varying data cases diminish its applicability in more intricate settings.

Why XGboost Algorithm performed best & Naïve Biase performed worst in dataset of Phishing:

### *Performance Difference:*

XGBoost vs. Naive Bayes in Phishing Detection

1. XGBoost: A gradient boosting algorithm that builds multiple decision trees to secure difficult patterns in high-dimensional data like phishing URLs. It achieved 97% correctness due to its ability to manage non-linear relationships.
2. Naive Bayes: Guesses feature independence, which is often unlikely in phishing detection. Interrelated features like URL length and domain attributes lower its performance, resulting in 68% accuracy.

Analysis: XGBoost beats Naive Bayes by effectively portraying complicated attribute reliance, while Naive Bayes is reduced by its independence assumption.

### *Results of ML Algorithms*

**Table 1: Results of ML Algorithms**

| **Algorithms** | **Train Accuracy** | **Test Accuracy** | **Precision** | **Recall** | **F1-score** | **Runtime (seconds)** | **Confusion Matrix** |
|---|---|---|---|---|---|---|---|
| Random Forest | 100.00% | 96.72% | 97.05% | 96.28% | 96.67% | 1.6 | [[1126 31]<br>[ 45 1084]] |
| Decision Trees | 100.00% | 93.26% | 92.95% | 93.45% | 93.20% | 0.2 | [[1078 79]<br>[ 70 1059]] |
| Support Vector Machine (SVM) | 97.03% | 96.33% | 96.44% | 96.10% | 96.27% | 5.83 | [[1117 40]<br>[ 44 1085]] |
| Naive Bayes | 67.77% | 68.07% | 93.46% | 38.00% | 54.03% | 0.03 | [[1127 30]<br>[ 700 429]] |
| Logistic Regression | 94.43% | 95.71% | 95.90% | 95.39% | 95.65% | 0.1 | [[1111 46]<br>[ 52 1077]] |
| K-Nearest Neighbors (KNN) | 95.63% | 94.49% | 96.14% | 92.56% | 94.31% | 2.01 | [[1115 42]<br>[ 84 1045]] |
| XGBoost | 100.00% | **97.33%** | 97.42% | 97.17% | 97.29% | 0.52 | [[1128 29]<br>[ 32 1097]] |
| Boosted Decision Tree | 96.45% | 95.93% | 95.84% | 95.93% | 95.88% | 4.05 | [[1110 47]<br>[ 46 1083]] |
| AdaBoost | 94.90% | 95.36% | 95.31% | 95.31% | 95.31% | 1.12 | [[1104 53]<br>[ 53 1076]] |
| Extra Trees | 100.00% | 96.85% | 97.74% | 95.84% | 96.78% | 1.5 | [[1131 26]<br>[ 46 1083]] |
| LightGBM | 99.80% | 97.16% | 97.16% | 97.08% | 97.12% | 0.94 | [[1125 32]<br>[ 33 1096]] |
| CatBoost | 99.08% | 97.11% | 97.41% | 96.72% | 97.07% | 9.06 | [[1128 29]<br>[ 37 1092]] |

In the comparative analysis (Table 1), the details of the study samples analyzed through various machine learning algorithms for identifying the incidence of phishing attacks are presented and evaluated in this section.

XGBoost performs well when evaluated because boosting technique is robust that enhances the ability to rectify mistakes while has a unique way of handling overfitting by means of regularization. The positivity of it is that it can optimize its performance and features to interact with other features making it best suited for large datasets. Naive Bayes have particularly poor results because it assumes that all the features are independent, which is rarely the case, leading to inferior results again. Additionally, its basic structure and ability to work well with continuous inputs and varying data cases diminish its applicability in more intricate settings.

## CONCLUSION

To formulate new anti-phishing tactics, phishers refer to previous attempts aimed at countering the phenomenon. We thus require from a robust security that stays ahead a score of them. That is why early prevention by using machine learning models is urgently necessary. In contrast, any system that is built with principles derived from machine learning will have the statistical ability improve from each task that has been solved. In this work, we discussed and presented a comprehensive evaluation of machine learning approaches for phishing mitigation. From this proposal, a clear framework will be laid to facilitate the conduct of a comprehensive research on how application of advanced machine-learning techniques can be applied to enhance the situation regarding security even in a second of time. Like XGBoost algorithm, took 0.58 second to calculate the accuracy of 97.33% of test accuracy. So, we can see how robust are these algorithms day by day to detect the new techniques of phishing that phishers used smartly to steal credentials. The findings of the study suggest that because of the evolving nature of machine learning techniques, it is possible to prevent phishing that is, win the battle against phishers in the near future.

## REFERENCE

[1] Ramana, A.V., Rao, K.L., Rao, R.S., (2021). Stop-Phish: an intelligent phishing detection method using feature selection ensemble. Social Network Anal. Mining 11 (1), 1–9. https://doi.org/10.1007/s13278-021-00829-w.

[2] Anti-Phishing Working Group (2023). Phishing activity trends report. https://docs.apwg.org/reports/apwg_trends_report_q4_2023.pdf

[3] K S, Jishnu & Arthi, B. (2023). Review of the effectiveness of machine learning based phishing prevention systems. 050006. 10.1063/5.0175593.

[4] Sameen, Maria, Kyunghyun Han, and Seong Oun Hwang. "PhishHaven—an efficient real-time ai phishing URLs detection system." IEEE Access 8 (2020): 83425-83443.

[5] Benavides, E., Fuertes, W., Sanchez, S., & Sanchez, M. (2019). Classification of phishing attack solutions by employing Deep learning techniques: A Systematic literature review. In Smart innovation, systems and technologies (pp. 51–64). https://doi.org/10.1007/978-981-13-9155-2_5

[6] Asif, A.U.Z., Shirazi, H., Ray, I. (2023). Machine Learning-Based Phishing Detection Using URL Features: A Comprehensive Review. In: Dolev, S., Schieber, B. (eds) Stabilization, Safety, and Security of Distributed Systems. SSS 2023. Lecture Notes in Computer Science, vol 14310. Springer, Cham. https://doi.org/10.1007/978-3-031-44274-2_36

[7] Alnemari, S., & Alshammari, M. (2023). Detecting phishing domains using machine learning. Applied Sciences, 13(8), 4649. https://doi.org/10.3390/app13084649

[8] Arshad, A, Rehman, A.U., Javaid, S., Ali, T.M., Sheikh, J.A., Azeem, M., 2021. A Systematic Literature Review on Phishing and Anti-Phishing Techniques. arXiv. https://doi.org/10.48550/arXiv.2104.01255.

[9] Safi, A., & Singh, S. (2023). A systematic literature review on phishing website detection techniques. Journal of King Saud University. Computer and Information Sciences/Maǧalaẗ Ǧam'aẗ Al-malīk Saud: Ùlm Al-ḥasib Wa Al-ma'lumat, 35(2), 590–611. https://doi.org/10.1016/j.jksuci.2023.01.004

[10] Yang, R., Zheng, K., Wu, B., Wu, C., & Wang, X. (2021). Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning. Sensors, 21(24), 8281. https://doi.org/10.3390/s2124828

[11] Mosa, D. T., Shams, M. Y., Abohany, A. A., El-Kenawy, E. M., & Thabet, M. (2023). Machine learning techniques for detecting phishing URL attacks. Computers, Materials & Continua/Computers, Materials & Continua (Print), 75(1),1271–1290. https://doi.org/10.32604/cmc.2023.036422

[12] Ojewumi, T. O., Ogunleye, G. O., Oguntunde, B. O., Folorunsho, O., Fashoto, S. G., & Ogbu, N. (2022). Performance evaluation of machine learning tools for detection of phishing attacks on web pages. Scientific African https://doi.org/10.1016/j.sciaf.2022.e01165.

[13] Hannousse, A., & Yahiouche, S. (2021). Web page phishing detection. Data.mendeley.com, 3. https://doi.org/10.17632/c2gw7fy2j4.3 (Source paper)

[14] Awasthi, A., & Goel, N. (2022). Phishing website prediction using base and ensemble classifier techniques with cross-validation. Cybersecurity, 5(1). https://doi.org/10.1186/s42400-022-00126-9

[15] Aljofey, A., Jiang, Q., Rasool, A., Chen, H., Liu, W., Qu, Q., & Wang, Y. (2022). An effective detection approach for phishing websites using URL and HTML features. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-10841-5

[16] Aung, E.S., Zan, C.T., & Yamana, H. (2019). A Survey of URL-based Phishing Detection.

[17] Alsariera, Y. A., Adeyemo, V. E., Balogun, A. O., & Alazzawi, A. K. (2020). AI Meta-Learners and Extra-Trees algorithm for the detection of phishing websites. IEEE Access, 8, 142532–142542. https://doi.org/10.1109/access.2020.3013699

[18] Jampen, Daniel & Gür, Gürkan & Sutter, Thomas & Tellenbach, Bernhard. (2020). Don't click towards an effective anti-phishing training. A comparative literature reviews. Human-centric Computing and Information Sciences. 10. 10.1186/s13673-020-00237-7.

[19] hobrekar, L., Munshi, Q., & Naik, S. (2022). Machine Learning in Policing Counterfeit Websites. International Research Journal of Innovations in Engineering and Technology, 06(01), 46–53. https://doi.org/10.47001/irjiet/2022.601010

[20] Pujara, P., & Chaudhari, M. B. (2018). Phishing website detection using machine learning: a review. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(7), 395-399.

[21] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. Expert Systems with Applications, 117, 345-357.

[22] Shabu, S. J., Refonaa, J., & Dhamodaran, S. (2022). Phishing Website Detection using Machine Learning Algorithm. Mathematical Statistician and Engineering Applications, 71(3s2), 579-587

[23] Shikalgar, S., Sawarkar, S. D., & Narwane, S. (2019). Detection of URL based phishing attacks using machine learning. Int. J. Eng. Res. Tech. (IJERT), 8(11).

[24] Su, Y. (2020, June). Research on website phishing detection based on LSTM RNN. In 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (Vol. 1, pp. 284-288). IEEE.

[25] A.A, A., & K, P. (2020, June 1). Towards the Detection of Phishing Attacks. IEEE Xplore. https://doi.org/10.1109/ICOEI48184.2020.9142967

[26] Sindhu, S., Patil, S.P., Sreevalsan, A., Rahman, F., Saritha, A.N., 2020. Phishing detection using random forest, SVM and neural network with backpropagation. In: Proceedings of the International Conference on Smart Technologies in Computing, Electrical and Electronics, ICSTCEE 2020, pp. 391–394. https://doi.org/10.1109/ICSTCEE49637.2020.9277 256.

[27] Benavides-Astudillo, E., Fuertes, W., Sanchez-Gordon, S., Nuñez-Agurto, D., & Rodríguez-Galán, G. (2023). A phishing-attack-detection model using Natural Language Processing and Deep Learning. Applied Sciences (Basel, Switzerland), 13(9), 5275. https://doi.org/10.3390/app13095275

[28] Shahrivari, V., Darabi, M. M., & Izadi, M. (2020). Phishing detection using Machine Learning techniques. Retrieved from http://arxiv.org/abs/2009.11116

[29] Butt, U.A., Amin, R., Aldabbas, H. et al. Cloud-based email phishing attack using machine and deep learning algorithm. Complex Intell. Syst. 9, 3043–3070 (2023). https://doi.org/10.1007/s40747-022-00760-3

[30] Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A., & Elsoud, E. A. (2022). An intelligent cyber security phishing detection system using deep learning techniques. Cluster Computing. https://doi.org/10.1007/s10586-022-03604-4

[31] Singh, S., Beniwal, H., 2021. A survey on near-human conversational agents. J. King Saud Univ. – Comput. Inf. Sci. Volume 34 (10, Part A), 8852–8866. https://doi.org/10.1016/j.jksuci.2021.10.013.

[32] Alkawaz, Mohammed & Steven, Stephanie & Hajamydeen, Asif Iqbal & Ramli, Rusyaizila. (2021). A Comprehensive Survey on Identification and Analysis of Phishing Website based on Machine Learning Methods. 82-87. 10.1109/ISCAIE51753.2021.9431794.

[33]. Butnaru, A., Mylonas, A., & Pitropakis, N. (2021). Towards Lightweight URL-Based Phishing Detection. Future Internet, 13(6), 154. https://doi.org/10.3390/fi13060154.

[34]. Ozker, Ugur & Sahingoz, Ozgur. (2020). Content Based Phishing Detection with Machine Learning. 1-6. 10.1109/ICEE49691.2020.9249892.

[35]. Geyik, B., Erensoy, K., Kocyigit, E., 2021. Detection of Phishing Websites from URLs by using Classification Techniques on WEKA. In: Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021, pp. 120–125. https://doi.org/10.1109/ICICT50816.2021.9358642.

[36] Jain, A.K., Gupta, B.B., 2019. A machine learning based approach for phishing detection using hyperlinks information. J. Ambient Intell. Hum. Comput. 10 (5),2015–2028. https://doi.org/10.1007/s12652-018-0798-z.

[37] Basit, A., Zafar, M., Javed, A.R., Jalil, Z., 2020. A Novel Ensemble Machine Learning Method to Detect Phishing Attack. In: Proceedings - 2020 23rd IEEE International Multi-Topic Conference INMIC 2020. https://doi.org/10.1109/INMIC50486.2020.931821 0.

[38] Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., Shukla, S., 2022. Applications of Deep Learning for Phishing Detection: A Systematic Literature Review Knowl. Inf.Syst. 64 (6). https://doi.org/10.1007/s10115-022-01672-x.

[39] Zhu, E., Ju, Y., Chen, Z., Liu, F., & Fang, X. (2020). DTOF-ANN: An Artificial Neural Network phishing detection model based on Decision Tree and Optimal Features. Applied Soft Computing, 95, 106505. https://doi.org/10.1016/j.asoc.2020.106505.

[40]. Korkmaz, M., Kocyigit, E., Sahingoz, O.K., Diri, B., 2021. Phishing Web Page Detection Using N-gram Features Extracted from URLs. In: HORA 2021 - 3rd International Congress on Human-Computer Interaction, Optimization and Robotic. Applications, Proceedings. https://doi.org/10.1109/HORA52670.2021.946137 8.

[41] Van Dooremaal, B., Burda, P., Allodi, L., & Zannone, N. (2021, August). Combining text and visual features to improve the identification of cloned webpages for early phishing detection. In Proceedings of the 16th International Conference on Availability, Reliability and Security (pp. 1-10).

[42]. B. Riskhan et al., "An Adaptive Distributed Denial of Service Attack Prevention Technique in a Distributed Environment," Sensors, vol. 23, no. 14, p. 6574, 2023.

[43] H. A. J. Fufan, B. Riskhan, K. Hussain, H. Bin Jazri, S. H. Alajmani, and K. Sharma, “SYN Flood Attack Detection and Prevention Technique (SFaDMT) in Distributed Environment,” in 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), IEEE, 2024, pp. 1–9.

[44] B. Rishkan, M. H. Usmani, M. A. U. Sheikh, H. Bin Jazri, N. A. Khan, and K. Sharma, “Detection of SYN Flood Attack Based on Bays Estimator (DsFaBe) in MANET,” in 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), IEEE, 2024, pp. 1–5.