e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

DETECTION OF PHISHING ATTACKS USING MACHINE LEARNING TECHNIQUES

Samaila Kasimu Ahmad*1, Babagana Ali Dapshima*2, Yasmin Chuupa Essa*3

*1Department Of Computer Science And Application, Sharda University, Greater Noida UP India.

*2,3Department Of Computer Science And Engineering, Sharda University, Greater Noida UP India.

DOI: https://www.doi.org/10.56726/IRJMETS60054

ABSTRACT

The study evaluated several machine learning techniques for detecting phishing attacks, including Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR). Two datasets were used - one from PhishTank and another from the UCI machine learning repository. Results showed that the Random Forest model achieved the highest accuracy across multiple metrics. On the PhishTank dataset, RF had the best K-fold cross-validation accuracy at 99.55%, feature selection accuracy at 99.00%, and hyperparameter tuning accuracy at 99.45%. The XGBoost model performed well too, with 99.16% K-fold accuracy on PhishTank. On the UCI dataset, XGBoost had the highest K-fold accuracy at 97.16%, while RF still demonstrated maximum accuracy for feature selection and hyperparameter tuning. Logistic Regression consistently showed the lowest accuracy across datasets and metrics. The proposed approach was validated against other researchers' work on PhishTank, achieving 98.80% accuracy, which was compared favorably. ROC curves further illustrated the strong performance, especially for the top-performing models. The study demonstrated that using selected features and hyperparameter tuning could enhance detection accuracy. The machine learning algorithms, particularly Random Forest, outperformed other state-of-the-art techniques in accurately identifying phishing attacks. The high accuracy metrics indicate the proposed framework's effectiveness in detecting phishing attempts.

Keywords: Detection, Phishing Attack, Phish Tank, Hyper-Parameter, Machine Learning.

I. INTRODUCTION

Phishing continues to be a widespread and persistent cyber threat for valid reasons. For nearly three decades, it has proven to be a highly efficient method for breaching a company's defenses, primarily due to its adept manipulation of individuals. Despite increased awareness and dedicated resources aimed at preventing phishing, it remains as prevalent as ever [1].

When discussing phishing, it is crucial to understand that its effectiveness is not solely dependent on the credibility of the bait; it also heavily relies on the use of various social engineering techniques to entice individuals. Cybercriminals excel in the art of persuasion, and phishing methods constantly adapt to influence people's decisions. Ultimately, human cognition and behavior underpin the success of phishing [2] [3]. Criminals target businesses and organizations to gain access to sensitive information, which they then use against not only their immediate victims but also the customers and constituents of those victims. Phishing attempts often imitate trusted authorities and common websites in ways that may not raise suspicion among those who are not experienced in recognizing such scams. The attack typically succeeds once the attacker gains access, which is often unintentionally granted [4].

Efforts to defend against phishing through technology-based methods frequently fall short due to their limitations, and because the human factor cannot always be relied upon to provide sufficient support. User behavior-based protection and prevention strategies also tend to falter due to inadequate or infrequent training, an inability to measure the right metrics, and situations that put users in lose-lose scenarios [5].

In the present day, everyone is interconnected online, utilizing various hardware and software, and gradually linking up with all aspects of life. Currently, 16% of the global population is internet users. While the internet offers numerous advantages, its misuse can have severe consequences in terms of cybersecurity [6]. Malicious actors are present on the web, tricking users into trusting their fraudulent websites and guiding them towards actions that expose their information. The solution isn't to shun the internet altogether but to acquire knowledge about these threats and exercise caution to avoid falling prey to such attacks.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

Cyber-attacks are advancing in tandem with technological advancements [7]. Attackers are now capable of producing counterfeit websites that are increasingly challenging to differentiate from genuine ones [8] [9]. Individuals are easily misled by these counterfeit pages, and they shouldn't bear full responsibility if their understanding of cyber security is limited [10]. It would be unjust to anticipate users to distinguish these sites solely based on visual clues. Nevertheless, this lack of knowledge could ultimately make them vulnerable to social or economic harm [11].

Given the substantial impact of these outcomes, this research endeavor seeks to create a definitive solution for accurately distinguishing between phishing and legitimate websites, safeguarding users from potential exploitation [12]. Phishing occurrences are prevalent across numerous sectors, including Online Banking, E-Commerce, HR and finance, and Social Networking [13]. Although many existing techniques like blacklist-whitelist approaches offer some protection against such attacks, they cannot identify zero-day attacks [14].

Attacks using phishing utilize fake emails, texting inquiries, cellphone calls, as well as internet sites to deceive people into revealing confidential details and installing malware. Unauthorized use of identity, debit/credit card scams, attacks by ransomware, breaches of information, and significant financial damage can all result from this [15]. It stands as the primary form of social engineering, where perpetrators deceive and manipulate individuals to disclose vital information. These attacks exploit human vulnerabilities and employ pressure tactics. The attacker often poses as a trusted entity, creating urgency that prompts hasty actions. Hackers and fraudsters prefer these methods due to their cost-effectiveness compared to breaching computer systems.

The FBI highlights phishing emails as the leading vehicle for ransomware distribution by hackers. According to [16]. International Business Machine Cooperation's 2022, expense associated with a breach of information study shows that fraudulent activity (phishing) has risen from the fourth to the second most common cause of data breaches. Data breaches resulting from phishing incidents are particularly costly, with victims facing an average financial impact of $4.91 million. Thus, considering such effects, the research Detecting Phishing Attacks via Machine Learning Techniques became a topic of Interest.

Phishing is a cybersecurity tactic where malevolent individuals impersonate trusted sources in their messages. These misleading emails aim to trick users into carrying out measures such as loading malicious files, clicking on inappropriate links, or revealing sensitive data such as login credentials. A phishing attack is the most common form of online social engineering, a broad category encompassing efforts to manipulate or deceive computer users [17].

Types of Phishing

Email-based phishing, spear phishing, whaling, smishing, vishing, and phishing anglers are diverse forms of cyberattacks that exploit various communication channels to deceive victims. Email-based phishing involves mass emails with fake domain names to imitate legitimate organizations. Spear phishing targets specific individuals using detailed personal information to enhance email credibility, often leading victims to make financial transactions [18]. Whaling focuses on high-ranking executives, utilizing publicly available data to create sophisticated, personalized attacks. Smishing and vishing use phone-based methods, with smishing involving fake SMS messages and vishing employing fraudulent phone calls, where attackers pose as bank representatives to steal sensitive information. Phishing anglers exploit social media by creating fake profiles resembling real organizations, tricking users into sharing personal information or visiting malicious websites under the guise of customer support. These tactics highlight the evolving strategies cybercriminals use to exploit human vulnerabilities across different platforms [19].

Phishing Attack Process

The diagram below shows how phishing attack is carried out.

From figure 1, a phishing attack typically forms part of a broader campaign to ensnare numerous victims within a wide range of potential targets. From its initial inception to the successful acquisition of credentials, a phishing attack involves four distinct phases that must be carried out. We will now delve into each of these individual phases in the following graphic.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

1. In Phase 1: A malevolent hacker, posing as a reputable source, sends an email or message to the target, often requesting them to click on a third-party link under the guise of a security check or a simple software update.
2. In Phase 2: the target believes that the email originates from the mentioned sender, whether it's a bank or a company, and proceeds to click on the malicious link, leading them to a counterfeit webpage that closely mimics a legitimate site.
3. In Phase 3: while on the fraudulent website, the user is prompted to provide sensitive information such as account credentials for a specific online platform. Once this information is submitted, it is transmitted to the hacker responsible for creating the deceptive website and email.
4. Finally, in Phase 4: upon receiving the acquired account credentials, the hacker gains access to the user's account or may sell the gathered information to the highest bidder on the internet.

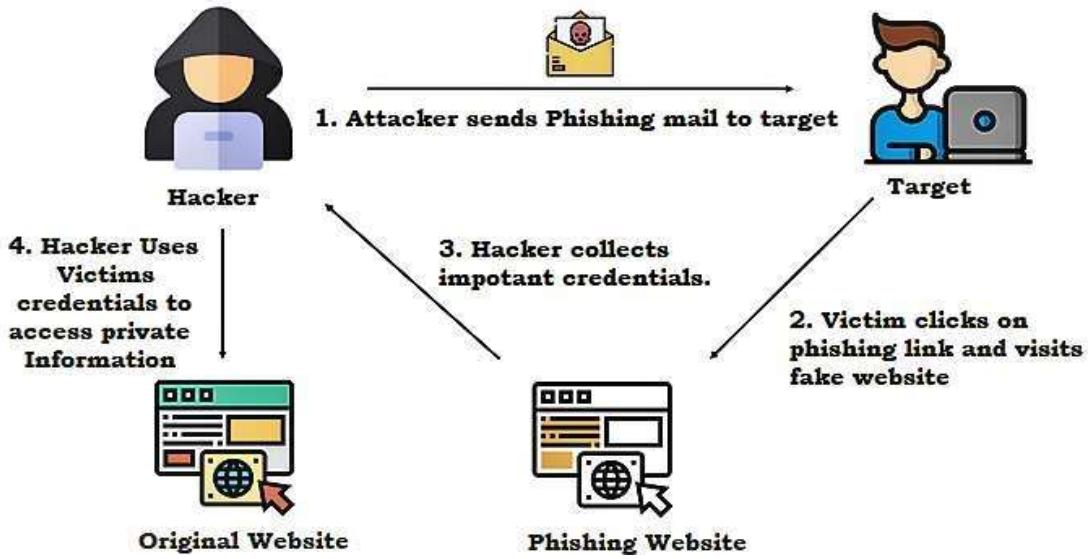

Figure 1: Phishing attack process

Indicators of Attacks by Phishing

Phishing emails often employ threats or a sense of urgency to prompt immediate action, exploiting hurried readings that may overlook inconsistencies [20]. Signs of phishing include unusual message styles, such as inappropriate language or tone, and requests for non-standard actions. Linguistic errors and discrepancies in web addresses are additional red flags [21]. Phishing emails may also direct recipients to fraudulent login pages to steal credentials or financial information. Organizations can combat phishing by educating employees on awareness, implementing email security solutions, and utilizing endpoint monitoring. Simulated phishing attack assessments help gauge the effectiveness of training, while granting access based on the principle of least privilege can protect sensitive data [22]. Check Point's email security system offers proactive protection against advanced phishing attacks, reducing organizational vulnerability.

Concept of Machine Learning

In the 1950s the concept of machine learning originated, The study of machine learning came into effect in the early 1990s [23]. It is a sub-core area of AI (Artificial Intelligence) that permits machines to access data and learn from the data which helps make work much easier for humans. The algorithms used by Machine learning (ML) is a computational method of directly studying information in a given data without depending on predetermined equations as it models [24] [25]. It's regarded as one of the vital approaches to Artificial Intelligence that shows some potential in various domains such as in medical, finance, business, and industry. ML is a modern statistical tool that aids in identifying the data output and their future use [26].

There are numerous advantages in the use of ML especially toward prediction and analysis of data. The architecture of machine learning explains the various steps involved in the transformation of unprocessed data to processed data (trained dataset) which allows decision-making in the system [27].

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

Machine learning is categories into three (3) types namely;

1. Supervised Machine Learning

This involves training a system with labeled data and the machine after adopting the training has the potential to predict the outcome based on the trained data, which is referred to as the supervised learning approach. It is a process involving a supervisor who guides and tells the machine what to measure and results are also expected. For instance, a teacher teaches a pupil about types of shapes like Circles, squares, triangles, etc. and after the pupil has learned, then asks them the name of the shapes of the same objects he taught [28]. In machine learning, the supervisor is always referred to as level data [29].

The below figure (Fig 1) shows the supervised learning approach.

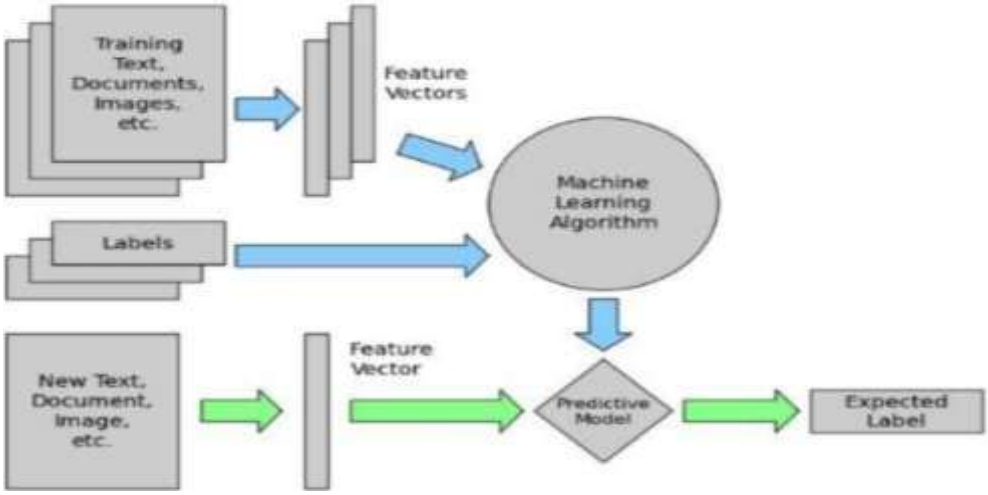

Figure 2: A model of supervised learning

2. Unsupervised Machine Learning

Unsupervised learning in artificial intelligence is a type of machine learning that learns from data without human supervision. Unlike supervised learning, unsupervised machine learning models are given unlabeled data and allowed to discover patterns and insights without any explicit guidance or instruction [30].

Whether you realize it or not, artificial intelligence and machine learning are impacting every aspect of daily life, helping to turn data into insights that can improve efficiencies, reduce costs, and better inform decision-making. Today, businesses are using machine learning algorithms to help power personalized recommendations, real-time translations, or even automatically generate text, images, and other types of content [31].

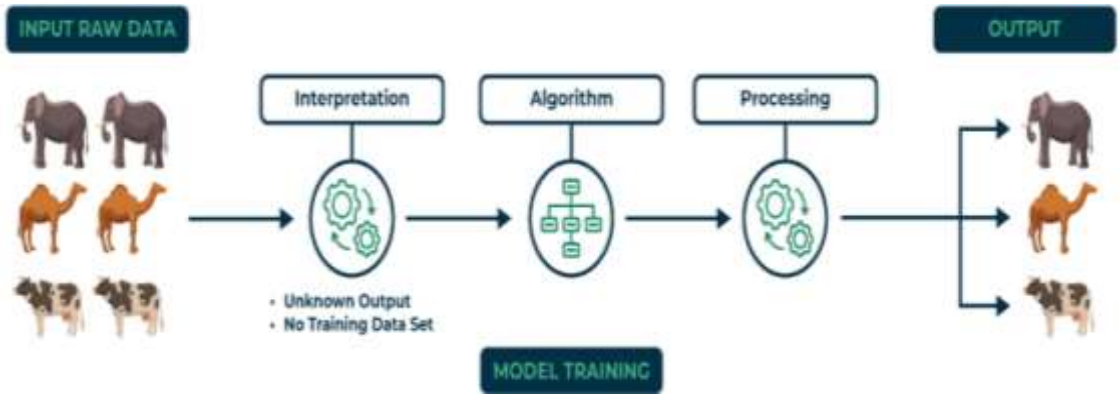

Figure 3: A model of unsupervised learning

3. RML (Reinforcement Machine learning)

RML belongs to the family of machine learning approaches which deal with outcomes based on feedback. The agents understand the environment by carrying out curtains action and also observe the results of the action they perform [32]. A positive response is dedicated to every good he does and vice-versa.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

Figure: 4: Reinforcement machine learning approach.

II. METHODOLOGY

When creating a method for detecting phishing websites, the methodology can differ depending on the particular system under development. You can see the recommended approach and its flow in Figure 5.

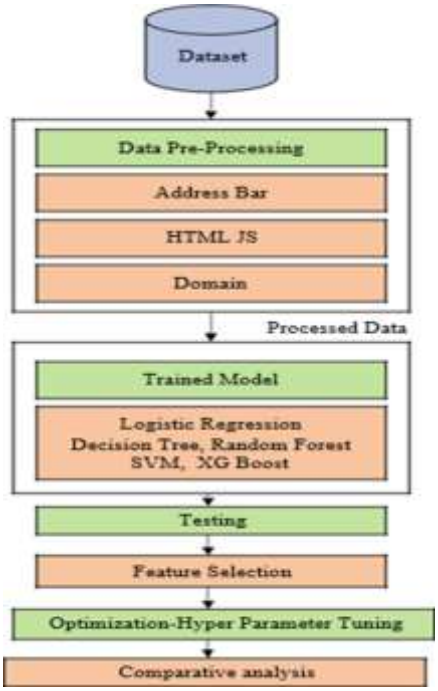

Figure 5: System flow for the proposed methodology

Implementation Methodology

The implementation methodology is explicitly discussed below;

First phase

In this phase, the dataset collected included both reputable websites and widely recognized phishing websites for training and testing the system.

Description of the Datasets

The first dataset, known as Phish Tank (Dataset 1), comprised 10,000 URLs equally distributed between two categories: 5,000 phishing URLs obtained from Phish Tank [] All sites used for phishing were labeled "1," whereas all authentic Sites were labeled "0." The coding language Python was used to extract features from this dataset, yielding twenty-five attributes taken from the mixed data set. This dataset was then used to train the model. The first tableau shows the retrieved characteristics and the corresponding information types.

The second set, (Dataset 2), was obtained from the UCI machine-learning storage facility and contained eleven thousand and fifty-five URLs, 6,157 of which were phishing samples and 4,898 of which were legal.

Second phase

The second stage involved two essential stages: model training and cross-validation. In the model training phase, the system underwent training using machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Extreme Gradient Boosting, utilizing extracted features and

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

labeled data. The model's performance was assessed through K-fold cross-validation, which divided the data into ten subsets for evaluation.

Third Phase

Table 1: Description of features

S/N.	Feature Description	Data Type
1	Domain of URL	E
2	Presence of IP Address	G
3	Presence of "@" Symbol	G
4	URL Length	G
5	URL Depth	F
6	Embedded Domain	G
7	Create HTTP	G
8	Services for Short URLs	G
9	Website address should be prefixed and suffixes	G
10	Server should store DNS records	G
11	Engagement of server	G
12	Antiquity the Domain	G
13	Domain Expiration	G
14	Redirecting Iframes	G
15	Bar status should be customized	G
16	Disable clicking of the cursor	G
17	Move to a different site	G
18	Google Index	G
19	C(%)	F
20	C (?)	F
21	C(-)	F
22	C(=)	F
23	C (.)	F
24	C(www)	F
25	Label	G

Note: E, F, and G represent Character, Integer, and Boolean respectively

III. MODELING AND ANALYSIS

Table 2: Model Comparison

	Phish Tank			ULC		
MODEL	**TRAINING**	**TESTING**	**K-FOLD**	**TESTING**	**TRAINING**	**K-FOLD**
SVM	97.55	98.40	97.40	94.99	94.87	93.95
XBoost	98.44	99.20	99.16	97.68	96.95	95.99
RF	98.93	99.31	97.55	99.00	96.74	97.16

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

DT	98.99	96.99	98.11	98.85	93.63	95.94
LR	94.01	94.77	94.22	93.44	91.79	91.99

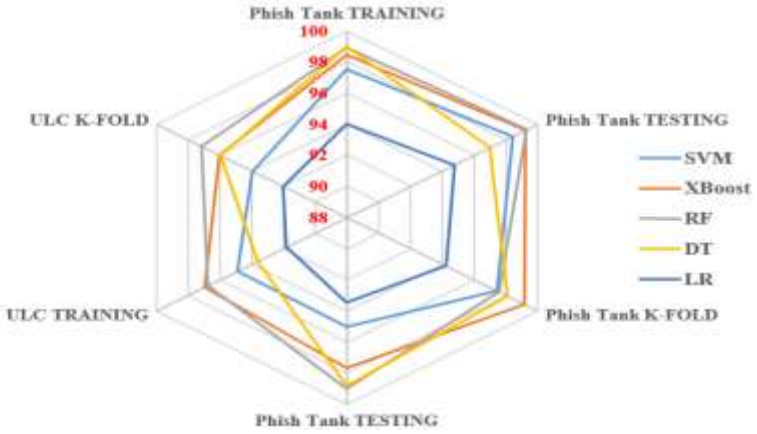

Figure 6: Model Comparison

Table 3: Hyper-parameter, feature selection, and K-fold Comparison

MODEL	K-FOLD	FEATURE SELECTION	HYPERPARAMETER TUNING
SVM	97.40	95.95	97.88
XBoost	99.16	97.98	98.74
RF	99.55	99.00	99.45
DT	97.11	98.11	98.00
LR	94.22	93.94	93.99

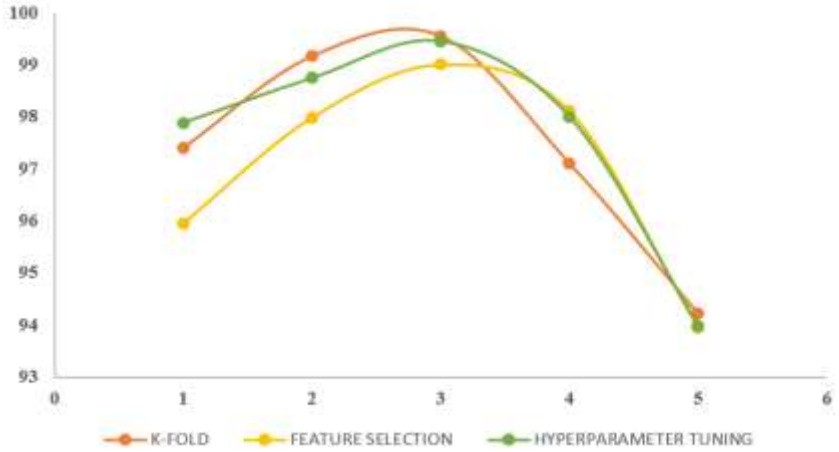

Figure 7: Hyper-parameter, feature selection, and K-fold Comparison

Table 4: Hyper-parameter, feature selection, and K-fold Comparison for UCI

MODEL	K-FOLD	FEATURE SELECTION	HYPERPARAMETER TUNING
SVM	95.00	95.95	96.99
XBoost	97.16	97.99	98.11
RF	97.11	98.15	98.50
DT	97.12	97.67	97.32
LR	93.53	93.66	93.80

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

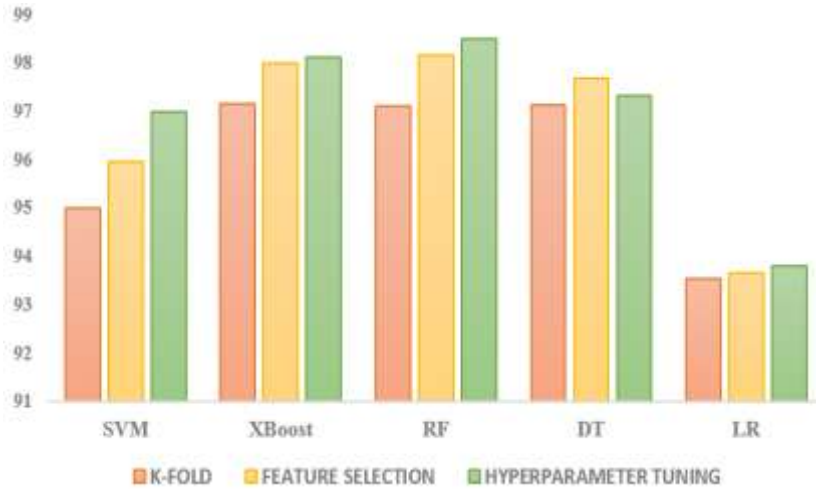

Figure 8: Graph for Hyper-parameter, feature selection, and K-fold Comparison for UCI

IV. RESULTS AND DISCUSSION

According to the approach outlined in phases 1 and 2, Table 2 displays an analysis of the two dataset's K-fold precision, assessment, as well as training results.

Table II reveals that the DT model achieved superior precision when working with PhishTank, while RF exhibited K-fold reliability and maximum performance using PhishTank. Notably, the LR model recorded the lowest K-fold accuracy at 93.94%. Additionally, Figure 2 illustrates the visual depiction of every single classification.

Feature selection represents a critical stage in the process. The recurrent element stripping technique was used to solve this. Moreover, hyperparameter adjustment was essential to the suggested methodology's optimization. The third table provides a full comparison of K-fold, choice of features, and hyperparameter modification using grid searching over PhishTank.

In the RF model, it is recorded as the most effective and precise strategy in tackling fraudulent activity (phishing attacks). It was observed at 99.55%, 99.00%, and 99.45% respectively. The LR model exhibited the lowest feature selection accuracy at 94%, as illustrated in Table 3. Table 4, shows that the XG boost model achieved the highest K-fold accuracy at 97.16%, while the RF classifier demonstrated the maximum accuracy in the remaining features. Conversely, the LR classifier had the lowest consistency.

Figures 8 and 9 depict receiver operator curve analysis for Phish Tank – ROC AUC curve. In Figure 10, shows the performance evaluation of the top model on the Phish Tank and UCI datasets across multiple metrics.

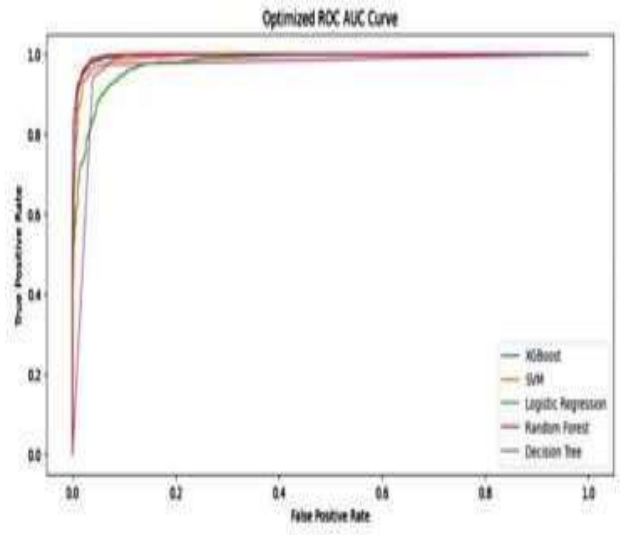

Figure 9: Phish Tank – ROC AUC curve.

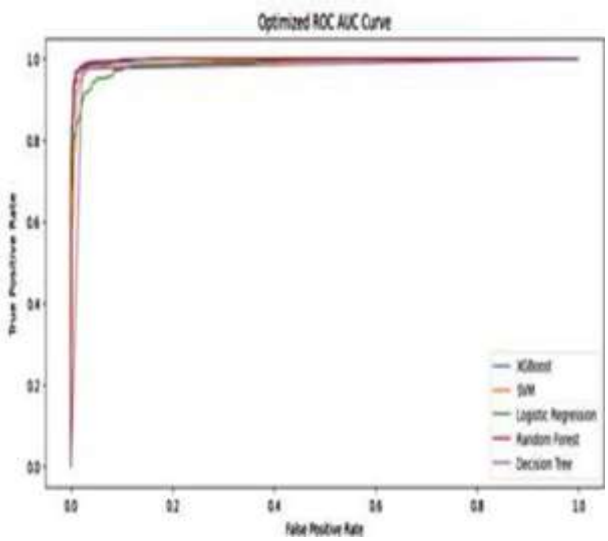

Figure 10: UCI – ROC AUC Curve

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

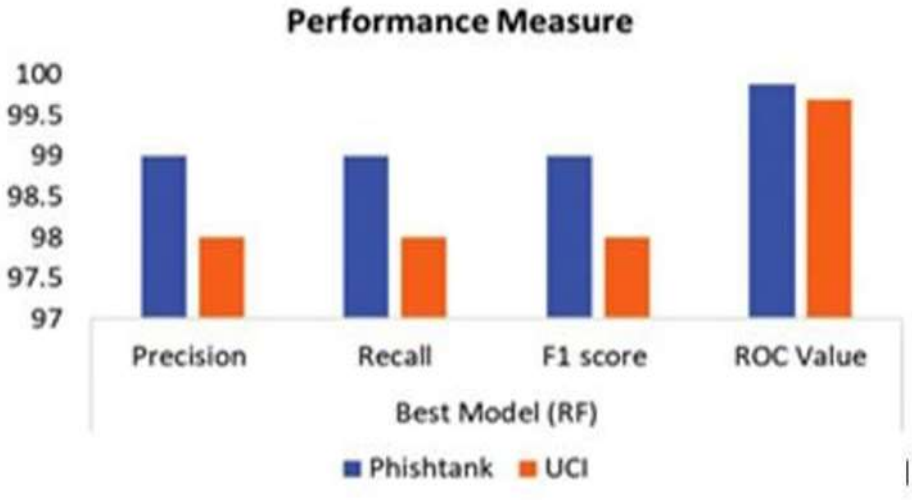

Figure 11: Comparative performance Analysis-Phish Tank and UCI dataset

Table 5: Phish Tank Validation

AUTHOR	TECHNIQUES	CONSISTENCY
1	Random Forest	87.30%
2	Random Forest	98.30%
3	RNN & CNN	90.25%
4	Random Forest	98.35%
5	Random Forest	99.10%

Validity of Approach

The validity of our suggested approach has been confirmed through collaboration with researchers currently active in the field, particularly those who have been involved in PhishTank projects. The outcomes of our proposed method, as indicated in Table 5, demonstrate its effectiveness, boasting an impressive accuracy rate of 98.80% when compared to the work of other researchers. Additionally, the approach's performance was assessed using the UCI dataset and verified through comparison with the work of established researchers. It is noteworthy that the methodology employed in our current research aligns closely with the approach of the RF classifier, and the results are consistent with those from prior research efforts.

V. CONCLUSION

Detecting phishing attempts with a high level of accuracy is crucial for a phishing website detection system. This study sought to detect phishing assaults by analyzing trends using automated learning techniques. In Fig. 9, the UCI ROC AUC curve illustrates our findings. It is observed that employing chosen features and hyper-parameter adjustment enhances the system's precision. The evaluation of machine learning algorithms surpassed other state-of-the- art techniques, yielding impressive accuracy results. The performance metrics within our proposed framework significantly contributed to the successful detection of phishing attacks.

VI. REFERENCES

[1] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade," IEEE Access, vol. 8, pp. 222310–222354, 2020, doi: 10. 1109/ACCESS. 2020. 3041951.

[2] E. Nowroozi, Abhishek, M. Mohammadi, and M. Conti, "An Adversarial Attack Analysis on Malicious Advertisement URL Detection Framework," IEEE Transactions on Network and Service Management, vol. 20, no. 2, pp. 1332–1344, Jun. 2023, doi: 10.1109/TNSM.2022.3225217.

[3] B. A. Dapshima, R. Mishra, and P. Tyagi, "Transformer faults identification via fuzzy logic approach," IJEECS, vol. 33, no. 3, p. 1327, Mar. 2024, doi: 10.11591/ijeecs.v33.i3.pp1327-1335.

[4] G. Mohamed, J. Visumathi, M. Mahdal, J. Anand, and M. Elangovan, "An Effective and Secure Mechanism for Phishing Attacks Using a Machine Learning Approach," Processes, vol. 10, no. 7, Art. no. 7, Jul. 2022, doi: 10.3390/pr10071356.

[5] A. Mughaid, S. AlZu'bi, A. Hnaif, S. Taamneh, A. Alnajjar, and E. A. Elsoud, "An intelligent cyber security phishing detection system using deep learning techniques," Cluster Comput, vol. 25, no. 6, pp. 3819–3828, Dec. 2022, doi: 10.1007/s10586-022-03604-4.

[6] U. A. Butt, R. Amin, H. Aldabbas, S. Mohan, B. Alouffi, and A. Ahmadian, "Cloud-based email phishing attack using machine and deep learning algorithm," Complex Intell. Syst., vol. 9, no. 3, pp. 3043–3070, Jun. 2023, doi: 10.1007/s40747-022-00760-3.

[7] Z. Azam, Md. M. Islam, and M. N. Huda, "Comparative Analysis of Intrusion Detection Systems and Machine Learning-Based Model Analysis Through Decision Tree," IEEE Access, vol. 11, pp. 80348–80391, 2023, doi: 10.1109/ACCESS.2023.3296444.

[8] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions," IEEE Access, vol. 10, pp. 36429–36463, 2022, doi: 10.1109/ACCESS.2022.3151903.

[9] B. A. Dapshima, "Constraints that Hinders Secure Software Implementation and Development Processes," IJRASET, vol. 12, no. 6, pp. 2400–2404, Jun. 2024, doi: 10.22214/ijraset.2024.63494.

[10] D. Rathee and S. Mann, "Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning," IJCA, vol. 183, no. 47, pp. 1–7, Jan. 2022, doi: 10.5120/ijca2022921868.

[11] L. Shalini, S. S. Manvi, N. C. Gowda, and K. N. Manasa, "Detection of Phishing Emails using Machine Learning and Deep Learning," in 2022 7th International Conference on Communication and Electronics Systems (ICCES), Jun. 2022, pp. 1237–1243. doi: 10.1109/ICCES54183.2022.9835846.

[12] F. S. Alsubaei, A. A. Almazroi, and N. Ayub, "Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics," IEEE Access, vol. 12, pp. 8373–8389, 2024, doi: 10.1109/ACCESS.2024.3351946.

[13] K. Joshi et al., "Machine-Learning Techniques for Predicting Phishing Attacks in Blockchain Networks: A Comparative Study," Algorithms, vol. 16, no. 8, Art. no. 8, Aug. 2023, doi: 10.3390/a16080366.

[14] A. Raza, K. Munir, M. S. Almutairi, and R. Sehar, "Novel Class Probability Features for Optimizing Network Attack Detection With Machine Learning," IEEE Access, vol. 11, pp. 98685–98694, 2023, doi: 10.1109/ACCESS.2023.3313596.

[15] T. O. Ojewumi, G. O. Ogunleye, B. O. Oguntunde, O. Folorunsho, S. G. Fashoto, and N. Ogbu, "Performance evaluation of machine learning tools for detection of phishing attacks on web pages," Scientific African, vol. 16, p. e01165, Jul. 2022, doi: 10.1016/j.sciaf.2022.e01165.

[16] M. Aljabri and S. Mirza, "Phishing Attacks Detection using Machine Learning and Deep Learning Models," in 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), Mar. 2022, pp. 175–180. doi: 10.1109/CDMA54072.2022.00034.

[17] A. Awasthi and N. Goel, "Phishing website prediction using base and ensemble classifier techniques with cross-validation," Cybersecurity, vol. 5, no. 1, p. 22, Nov. 2022, doi: 10.1186/s42400-022-00126-9.

[18] A. Basit, M. Zafar, A. R. Javed, and Z. Jalil, "A Novel Ensemble Machine Learning Method to Detect Phishing Attack," in 2020 IEEE 23rd International Multitopic Conference (INMIC), Nov. 2020, pp. 1–5. doi: 10.1109/INMIC50486.2020.9318210.

[19] A. Kovač, I. Dunđer, and S. Seljan, "An overview of machine learning algorithms for detecting phishing attacks on electronic messaging services," in 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), May 2022, pp. 954–961. doi: 10.23919/MIPRO55190.2022.9803517.

[20] N. F. Abedin, R. Bawm, T. Sarwar, M. Saifuddin, M. A. Rahman, and S. Hossain, "Phishing Attack Detection using Machine Learning Classification Techniques," in 2020 3rd International Conference on

Intelligent Sustainable Systems (ICISS), Dec. 2020, pp. 1125–1130. doi: 10. 1109/ICISS49785. 2020. 9315895.

[21] S. P. Ripa, F. Islam, and M. Arifuzzaman, "The Emergence Threat of Phishing Attack and The Detection Techniques Using Machine Learning Models," in 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Jul. 2021, pp. 1–6. doi: 10.1109/ACMI53878.2021.9528204.

[22] Z. B. Siddique, M. A. Khan, I. U. Din, A. Almogren, I. Mohiuddin, and S. Nazir, "Machine Learning-Based Detection of Spam Emails," Scientific Programming, vol. 2021, no. 1, p. 6508784, 2021, doi: 10. 1155/ 2021/6508784.

[23] S. Razaulla et al., "The Age of Ransomware: A Survey on the Evolution, Taxonomy, and Research Directions," IEEE Access, vol. 11, pp. 40698–40723, 2023, doi: 10.1109/ACCESS.2023.3268535.

[24] R. Zieni, L. Massari, and M. C. Calzarossa, "Phishing or Not Phishing? A Survey on the Detection of Phishing Websites," IEEE Access, vol. 11, pp. 18499–18519, 2023, doi: 10.1109/ACCESS.2023.3247135.

[25] V. Shahrivari, M. M. Darabi, and M. Izadi, "Phishing Detection Using Machine Learning Techniques." arXiv, Sep. 20, 2020. doi: 10.48550/arXiv.2009.11116.

[26] S. Naaz, "Detection of Phishing in Internet of Things Using Machine Learning Approach," IJDCF, vol. 13, no. 2, pp. 1–15, Mar. 2021, doi: 10.4018/IJDCF.2021030101.

[27] A. K. Dutta, "Detecting phishing websites using machine learning technique," PLOS ONE, vol. 16, no. 10, p. e0258361, Oct. 2021, doi: 10.1371/journal.pone.0258361.

[28] M. Somesha, A. R. Pais, R. S. Rao, and V. S. Rathour, "Efficient deep learning techniques for the detection of phishing websites," Sādhanā, vol. 45, no. 1, p. 165, Jun. 2020, doi: 10.1007/s12046-020-01392-4.

[29] A. Karim, S. Azam, B. Shanmugam, and K. Kannoorpatti, "Efficient Clustering of Emails Into Spam and Ham: The Foundational Study of a Comprehensive Unsupervised Framework," IEEE Access, vol. 8, pp. 154759–154788, 2020, doi: 10.1109/ACCESS.2020.3017082.

[30] D. M. Cao et al., "Advanced Cybercrime Detection: A Comprehensive Study on Supervised and Unsupervised Machine Learning Approaches Using Real-world Datasets," Journal of Computer Science and Technology Studies, vol. 6, no. 1, Art. no. 1, Jan. 2024, doi: 10.32996/jcsts.2024.6.1.5.

[31] V. K. Nadar, B. Patel, V. Devmane, and U. Bhave, "Detection of Phishing Websites Using Machine Learning Approach," in 2021 2nd Global Conference for Advancement in Technology (GCAT), Oct. 2021, pp. 1–8. doi: 10.1109/GCAT52182.2021.9587682.

[32] A. Alhogail and A. Alsabih, "Applying machine learning and natural language processing to detect phishing email," Computers & Security, vol. 110, p. 102414, Nov. 2021, doi: 10. 1016/j. cose. 2021. 102414.