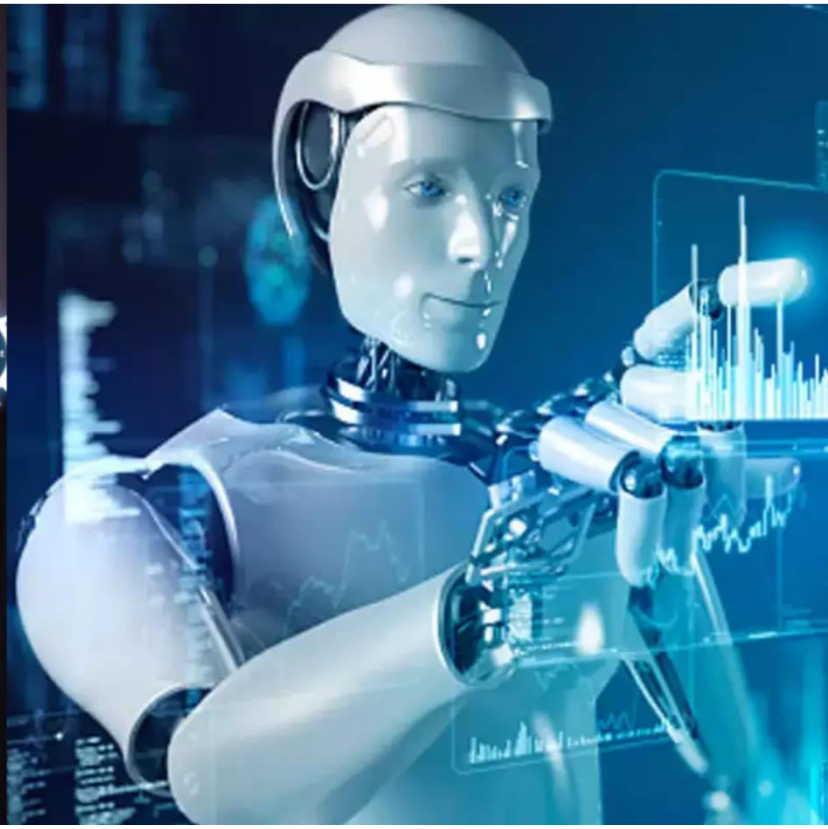
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

Intelligent Methods for Accurately Detecting Phishing Websites

Prof. Abhay Gaidhani, Sharvari Bhaladhare, Sakshi Deore, Madhura Pandit, Atharv Dharmadhikari

Professor, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, India

Student, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, India

Student, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, India

Student, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, India

Student, Department of Computer Engineering, Sandip Institute of Technology and Research Centre, Nashik, India

ABSTRACT: Phishing continues to pose a significant threat in the digital ecosystem, constantly evolving to exploit human behavior and technical vulnerabilities, thereby compromising sensitive information. To counter these threats, cybersecurity professionals have increasingly turned to machine learning (ML) techniques to automate the detection and mitigation of phishing attacks. This paper explores the use of the Gradient Boosting Classifier, which demonstrated strong performance with an accuracy of 97.4 percent and an F1 score of 97.7 percent when assessed alongside other ML models. The study conducts an in-depth analysis of various URL features to effectively distinguish between legitimate and malicious links. These features include structural attributes such as URL length, domain age, and the presence of subdomains, as well as lexical properties like textual content and obfuscation methods frequently employed in phishing schemes. In addition, semantic aspects are evaluated by analyzing web page content, SSL certificate details, and behavioural patterns linked to harmful domains. By leveraging this diverse set of characteristics, the machine learning model can achieve high precision and recall in phishing URL detection. The study also acknowledges several challenges, including class imbalance in datasets, complex feature engineering, and adversarial techniques used by sophisticated attackers to evade detection. Despite these obstacles, the results reinforce the effectiveness of machine learning, particularly Gradient Boosting, as a reliable approach for enhancing cybersecurity defences against phishing threats.

KEYWORDS: Cybersecurity Threats, Digital Security Solutions, Machine Learning Models, Pattern Recognition Techniques, Phishing Detection, URL Feature Analysis.

I. INTRODUCTION

Phishing attacks remain a persistent and rapidly evolving threat within the realm of cybersecurity. These attacks exploit user trust through deceptive tactics aimed at acquiring sensitive personal or financial information. Traditional detection methods, though once effective, often struggle to keep pace with the increasingly sophisticated techniques employed by modern cybercriminals. As a result, there has been a growing reliance on machine learning (ML) approaches by researchers and cybersecurity professionals to develop automated systems that can detect and mitigate phishing threats with greater efficiency and accuracy.

A crucial component of these systems involves the analysis of Uniform Resource Locators (URLs), which serve as gateways to online content. Since phishing often begins with deceptive URLs that mimic legitimate sources, analyzing URL characteristics becomes vital for early detection. Machine learning models can examine multiple dimensions of a URL, including its structural attributes such as length, number of subdomains, and domain age. Lexical features, such as the presence of suspicious keywords or obfuscation techniques, further help in distinguishing phishing attempts. In addition, semantic analysis evaluates contextual elements like webpage content, SSL certification, and domain behavior, enabling models to understand the intent behind the URLs more effectively.

Despite its promise, the application of machine learning in phishing detection is not without challenges. Issues such as dataset imbalance, the complexity of feature engineering, and concerns around data privacy and misuse present ongoing

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

hurdles. Furthermore, the dynamic and adversarial nature of phishing requires constant adaptation of ML models to counteract evolving attack strategies. Addressing these concerns calls for a multidisciplinary approach involving collaboration between academia, industry, and policy makers. Such collective efforts are essential for advancing ML-driven cybersecurity solutions capable of protecting digital infrastructures from phishing and other cyber threats.

II. LITERATURE SURVEY

In this paper, the authors tackle the growing threat of phishing attacks by leveraging both machine learning and deep learning techniques for effective detection of malicious URLs. They begin by introducing a feature engineering approach that extracts relevant characteristics from URLs to train machine learning models, achieving a commendable accuracy of 89.54% and an F1-score of 92.8%. Recognizing the limitations of manual feature extraction, the study further proposes deep learning models based on Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, which eliminate the need for handcrafted features by learning directly from raw URL data. Among these, the best-performing model achieved 98.1% accuracy and 98.7% F1-score, significantly surpassing the machine learning models. These findings highlight the superior capability of CNN and LSTM in capturing complex patterns in URLs, thereby offering a more robust and scalable solution for phishing detection [1].

In this paper, the authors present a systematic review and meta-analysis of existing user studies on phishing susceptibility, with a focus on understanding how demographic factors and training interventions influence users' ability to detect phishing attempts. While previous individual studies have explored the effects of variables such as age and gender, their findings have often been inconsistent, suggesting that older users are more susceptible, others indicating better performance among older participants, and many reporting no significant correlation at all. Similarly, studies on gender differences have yielded mixed outcomes. By synthesizing data from multiple studies across four major research databases, the authors aim to resolve these discrepancies and provide a more definitive evaluation. The meta-analysis reveals that age does have a significant impact on phishing susceptibility, with older users generally more vulnerable. Additionally, the analysis shows that females are more susceptible than males and that user training plays a critical role in enhancing phishing detection capabilities. These insights contribute valuable clarity to literature and support the development of more targeted and effective anti-phishing strategies [2].

In this paper, the authors conduct a comprehensive survey of recent advancements in phishing URL detection, emphasizing machine learning-based approaches as a means to combat the evolving tactics of attackers. Phishing, a prevalent form of social engineering, deceives users into revealing sensitive information through websites that mimic legitimate sources. Traditional detection methods, although available, are often time-consuming and ineffective against novel phishing techniques. To address this, researchers have increasingly turned to machine learning models that analyze various features such as URL structures, host information, and website content to differentiate malicious links from legitimate ones. The survey highlights how these models, when trained on carefully selected features, can significantly enhance detection accuracy and response time. By reviewing and comparing list-based and machine learning-based methods, this work provides insights into the current trends and challenges in the field, serving as a valuable resource for future developments in automated phishing detection [3].

In this paper, the authors present a survey of recent trends in phishing URL detection, with a particular focus on machine learning-based approaches aimed at addressing the growing sophistication of phishing attacks. Phishing, a form of social engineering, tricks users into disclosing sensitive information via deceptive websites that closely resemble legitimate sources. As attackers continuously develop new techniques, traditional detection methods—often reliant on blacklists or manual inspection—have become insufficient due to their time-consuming nature and limited adaptability. To overcome these limitations, researchers have increasingly employed machine learning models that utilize host-based features, URL characteristics, and website content to distinguish phishing sites from genuine ones. This survey reviews a range of list-based and machine learning-based techniques, highlighting their effectiveness, feature engineering strategies, and detection performance. The work offers valuable insights into current challenges and advancements, underscoring the potential of machine learning to enhance the accuracy and efficiency of phishing detection systems [4].

In this paper, the authors propose PhishHaven, an ensemble machine learning-based detection system designed to identify both human-crafted and AI-generated phishing URLs—addressing a growing concern sparked by systems

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

like DeepPhish, which use deep neural networks to create convincing phishing links. While prior research has explored machine learning and deep learning methods for detecting phishing attacks, most have focused solely on human-generated threats. PhishHaven distinguishes itself by incorporating advanced lexical analysis techniques and introducing novel strategies such as URL HTML Encoding for real-time classification and a URL Hit approach to resolve the persistent challenge of detecting shortened (tiny) URLs. The system further enhances accuracy and speed by utilizing a multi-threaded execution model and an unbiased voting mechanism for final classification decisions. Tested on a benchmark dataset of 100,000 URLs, PhishHaven achieves 98.00% accuracy, outperforming existing lexical-based detection systems. Theoretical analysis supports its robustness in identifying both current and future AI-generated phishing URLs, making it a significant advancement in phishing defense research [5].

In this paper, the authors present an intelligent phishing website detection system that functions as a browser extension, aiming to protect users from phishing attacks that exploit human vulnerabilities rather than software flaws. Recognizing the growing threat posed by phishing websites that deceive users into revealing sensitive information like usernames and passwords, the proposed system employs a supervised machine learning approach using the Random Forest algorithm, known for its robust classification performance. The authors emphasize selecting an optimal combination of features extracted from phishing websites to enhance the classifier's accuracy. Through careful feature analysis and model training, the system achieves an accuracy of 98.8% using 26 selected features. By integrating this high-performing classifier into a browser extension, the study demonstrates a practical and effective solution for real-time phishing detection, contributing to the broader effort of improving internet security through intelligent automation [6].

In this paper, the authors propose a model for detecting phishing websites using a URL-based detection approach powered by the Random Forest algorithm. Recognizing phishing as a fraudulent practice aimed at deceiving users to extract sensitive or financial information, the study addresses the increasing sophistication of phishing techniques that demand stronger security solutions. The proposed model is structured into three key phases: Parsing, Heuristic Classification, and Performance Analysis. Each phase utilizes distinct methods to process data effectively and enhance detection accuracy. By focusing on URL analysis and leveraging the robust classification capabilities of Random Forest, the model aims to improve the reliability and efficiency of phishing detection. This structured, multi-phase approach demonstrates a promising direction for strengthening online security against evolving phishing threats [7].

In this paper, the authors introduce a novel approach for detecting phishing attacks using natural language processing (NLP) techniques to analyze text and identify inappropriate statements that signal phishing attempts. Unlike traditional methods that primarily focus on technical features like URLs or website structures, this approach emphasizes semantic analysis of the text within phishing emails to detect malicious intent. By focusing on the natural language content of phishing attacks, the authors present a more human-centric detection model. The effectiveness of the proposed method is demonstrated through evaluation of a large benchmark set of phishing emails, showcasing its potential to enhance phishing detection by understanding the underlying malicious language patterns. This work represents a significant shift towards integrating linguistic analysis into cybersecurity solutions [8].

III. METHODOLOGY

The development of our phishing URL detection system follows a structured, step-by-step approach to ensure accuracy and resilience. We begin by loading a labelled dataset that contains various URL-based features along with their respective classifications. This data set forms the basis for our analytical and modelling efforts. Next, we perform exploration data analysis (EDA) to gain a deeper understanding of the data's structure and distribution. This step helps us identify significant trends, correlations, and irregularities that may influence model performance. Through EDA, we determine which features are likely to carry meaningful signals for phishing detection. To further enhance our understanding, we apply a range of data visualization techniques. We use tools such as histograms, box plots, and heatmaps to visualize feature interactions and their impact on the target variable. These visual insights guide us in refining our feature selection and engineering strategy. After preprocessing, we split the dataset into training and testing subsets. This division allows us to train models effectively while ensuring reliable performance evaluation on unseen data. We then experiment with multiple machine learning algorithms to develop predictive models, using performance metrics such as accuracy and F1 score to compare results.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

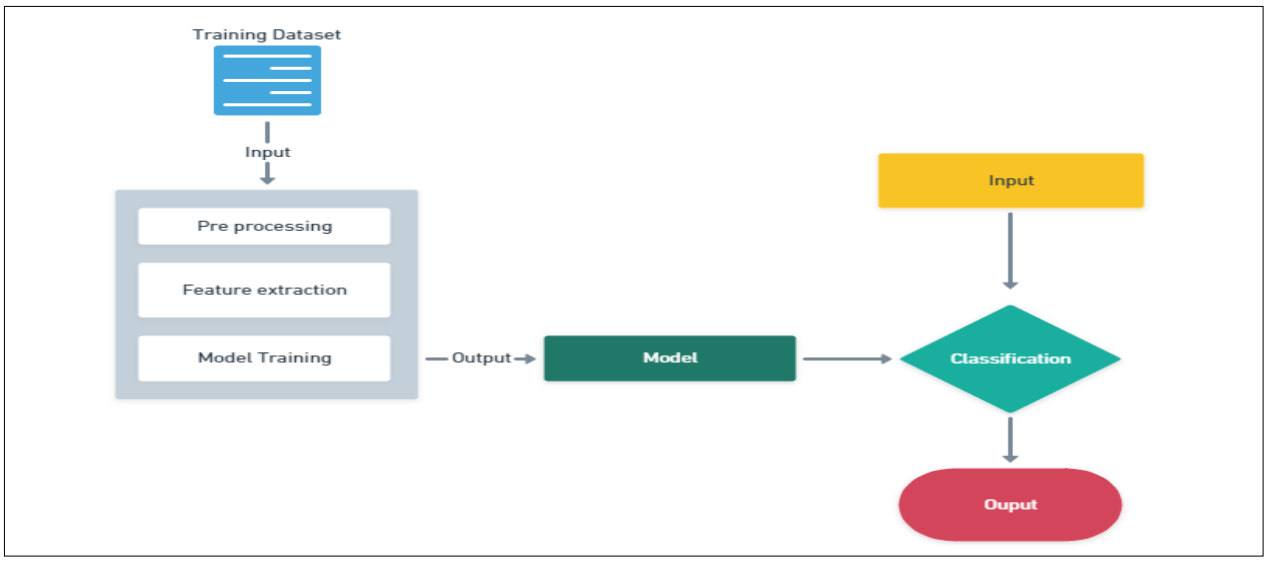


Figure 1 - System Architecture

Among all tested models, the Gradient Boosting Classifier demonstrates superior performance. It achieves the highest accuracy and F1 score, making it the most effective choice for our use case. We select this algorithm as the final model for deployment. To make the system accessible and user-friendly, we integrate the trained model into an interactive web application built with Streamlit. This interface enables users to input URLs and receive immediate classification results. Before prediction, the system automatically extracts relevant features from the URL to ensure consistent and comprehensive analysis.

Table 1 - Feature Extraction Techniques

Technique Used	Explanation
Structural Components	When building machine-learning models to detect phishing URLs, one of the most powerful sets of predictors comes from the URL's structural component features you can compute from the URL text, without external lookups. First, length-based features such as the total URL length, hostname length, path length, and query-string length can reveal obfuscation tactics: phishing URLs are often unusually long to conceal malicious tokens or parameters. Token-based features—counts of dots, hyphens, "@" symbols, query parameters, and the ratio of digits to letters—highlight unnatural patterns like multiple subdomain levels or autogenerated numeric domains. Special-character features, including the percentage of non-alphanumeric characters and the presence of suspicious keywords like "login," "verify," or "secure," flag attempts to lure users with reassuring language.
Lexical Elements	Lexical-element features for phishing-URL detection focus on three key areas: symbols, where we count occurrences of characters like @, #, ?, =, % and flag rare or excessive symbols (_, +, *, ;, ,, extra //) that phishers use to obfuscate or redirect; prefix/suffix patterns, such as hyphens at the start or end of domains (-bank.com, bank-.net), misplaced "www" (e.g. secure.www.bank.com.attacker.com), deceptive path segments beginning with or ending in tokens like login, verify, .php, .exe, and common TLDs used incorrectly as affixes (com-bank-login.com); and email-related indicators, including mailto: usage, counts of "email," "contact," or "support" tokens in paths or queries, "@" symbols outside the hostname, and full email-address patterns embedded in URLs—all of which betray attempts to harvest credentials or feign legitimacy.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

Semantic Attributes	Semantic-attribute features assess a URL's trustworthiness by combining protocol security, domain history, DNS presence, and web-visibility metrics: for example, HTTPS usage is a binary flag indicating whether the URL employs TLS (HTTPS sites are generally more trustworthy); domain registration duration and domain age (time since WHOIS creation) capture phishers' tendency to use newly registered or short-lived domains; DNS record availability (e.g., valid A, MX, NS records) checks that the domain resolves properly rather than pointing to parked or spoofed services; website traffic metrics (daily/monthly unique visitors from analytics providers) and PageRank values (or similar link-authority scores) measure popularity and link-based trust; Google indexing status (whether the URL appears in Google's index) indicates if the page is recognized by a major search engine; and the number of inbound links (backlinks from reputable sites) reflects external endorsement—together, these semantic attributes paint a holistic picture of a URL's legitimacy and visibility beyond mere syntactic patterns..

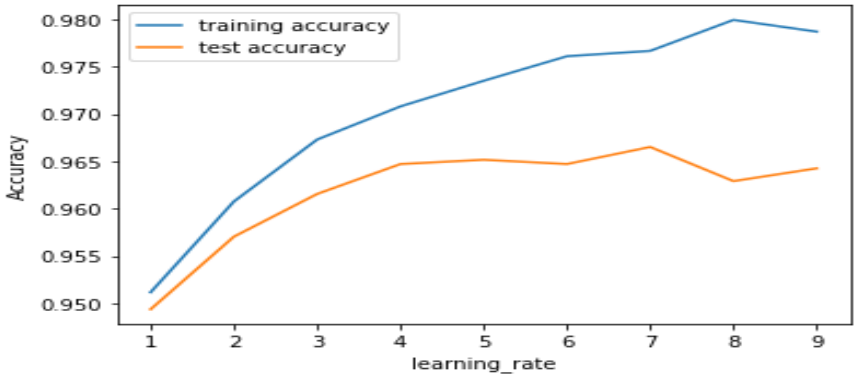
IV. ALGORITHM

Gradient Boosting is a powerful machine learning algorithm that builds an ensemble of weak learners, typically decision trees, in a sequential manner, with each new tree aiming to correct the errors made by the previous ones. The algorithm proceeds through the following key steps:

1. Start with an initial prediction, typically the mean value of the target variable for regression tasks or the log-odds for classification tasks.
2. In the first iteration, calculate the residuals, which are the errors between the actual values and the predicted values from the initial model.
3. Fit a Decision Tree to Residuals: Fit a weak learner (a decision tree) to these residuals. The goal of this tree is to predict the errors made by the previous model.
4. The model is updated by adding the predictions of the new tree, scaled by a learning rate, which controls how much the new model should influence the current model.
5. Repeat steps 2 to 4 for a predefined number of iterations or until the model's performance stops improving. Each new tree corrects the errors of the combined previous trees, with the model becoming more accurate with each iteration.
6. After completing all iterations, the final model is a weighted sum of the predictions of all the individual trees.

Throughout this process, the algorithm minimizes the loss function, which is a measure of how well the model fits the data. For regression tasks, the squared error is typically used, while for classification tasks, a log-loss function may be employed. The key strength of Gradient Boosting lies in its ability to sequentially improve upon weak models by focusing on correcting their mistakes, making it a highly effective method for both regression and classification tasks. However, careful hyperparameter tuning, particularly of the learning rate, tree depth, and the number of iterations, is necessary to avoid overfitting and to ensure the model generalizes well to unseen data.

V. RESULT AND ANALYSIS



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

These functions delve into the semantic attributes of URLs by assessing a wide range of indicators that help evaluate the legitimacy and trustworthiness of a web resource. One of the primary aspects examined is the use of HTTPS, which is a critical component for ensuring secure data transmission between users and websites. The presence of HTTPS in a URL often signifies that the website takes security seriously and adheres to encryption standards. In addition to HTTPS, the functions also analyze the domain registration period and domain age, which reflect how long a domain has been active. Typically, phishing websites have a short lifespan and are newly registered, while legitimate websites tend to have longer domain histories. The availability and integrity of DNS records are also reviewed, as the absence or inconsistency of DNS records can be a red flag indicating potential malicious behavior. By incorporating these elements into the evaluation process, the system gains a deeper understanding of a website's identity and operational background.

Beyond security and registration-related indicators, semantic analysis extends to factors that reflect a website's credibility and influence within the digital ecosystem. These include traffic volume, which shows how frequently a website is accessed by users, and PageRank, which measures the importance of web pages based on the number and quality of links they receive from other sites. High traffic and PageRank values often indicate a website's relevance and reliability. The system also checks whether a website is indexed by Google, as most legitimate websites are regularly crawled and listed by search engines. Furthermore, the number of external links pointing to the URL is analyzed, providing insight into how interconnected the site is within the web. A larger number of reputable inbound links can serve as a strong indicator of authenticity and trust. Together, these semantic features play a pivotal role in assessing the overall legitimacy, visibility, and authority of a URL, thereby strengthening the accuracy of phishing detection mechanisms.

Table 2. Accuracy of various model

No.	ML Model	Accuracy	F1 Score	Recall	Precision
1	Gradient Boosting Classifier	97.4%	97.7%	99.4%	98.6%
2	Random Forest	96.7%	99.3%	99.3%	99.0%
3	Support Vector Machine	96.4%	96.8%	98.0%	96.5%
4	Naive Bayes Classifier	60.5%	45.4%	29.2%	99.7%

Based on the evaluation metrics, none of the other machine learning models demonstrate performance levels comparable to those achieved by the Gradient Boosting Classifier. Across all key indicators, including accuracy, F1-score, recall, and precision, the Gradient Boosting Classifier consistently outperforms the other algorithms. This superior performance underscores its effectiveness in reliably identifying phishing URLs with a higher degree of accuracy. One possible reason for this enhanced capability is the model's iterative learning approach, which allows it to progressively correct errors by focusing more on difficult-to-classify instances. This adaptive mechanism significantly contributes to its robustness and precision in handling complex classification tasks, particularly in the context of phishing detection.

VI. CONCLUSION

We develop and evaluate a Gradient Boosting Classifier model specifically for the detection of phishing URLs. The model demonstrates consistently strong performance across multiple evaluation metrics, including accuracy, F1 score, recall, and precision, on various train-test splits. Notably, the model achieves an impressive accuracy of 98.9% on the training data and 97.4% on the testing data, which underscores its robustness in generalizing across different data samples. The F1 score, which provides a balanced measure by accounting for both precision and recall, reaches 99.0% for training and 97.7% for testing. This indicates that the model effectively maintains a balance between true positive detection and minimizing false alarms. Furthermore, the recall score, a key metric in phishing detection since it reflects the model’s ability to identify all actual phishing instances, stands at 99.4% and 98.9% for training and testing respectively, demonstrating the model’s high sensitivity to malicious URLs.

The model has demonstrated excellent performance, with a high precision score of 98%, indicating its strong ability to correctly identify phishing URLs. During the testing phase, the detection reliability has remained consistently high, ranging from 96% to 98.6%, showcasing its effectiveness. Moving forward, there are opportunities to further enhance

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

the model by fine-tuning hyperparameters, incorporating additional feature engineering techniques, and leveraging ensemble methods to improve accuracy even more. Additionally, exploring feature importance analysis could help pinpoint the key attributes driving the classification, leading to a more interpretable and optimized phishing detection system that is well-suited for real-world applications

REFERENCES

[1] R. Ferdaws and N. E. Majd, "Phishing URL Detection Using Machine Learning and Deep Learning," 2024 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2024, pp. 0485-0490, doi: 10.1109/AIIoT61789.2024.10579005.

[2] S. Baki and R. M. Verma, "Sixteen Years of Phishing User Studies: What Have We Learned?," in IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 2, pp. 1200-1212, 1 March-April 2023, doi: 10.1109/TDSC.2022.3151103.

[3] T. Shrivastava, S. Bhatt, N. R. Roy and A. Bhasney, "Phishing URL Detection Using Machine Learning: A Survey," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 1992-1997, doi: 10.1109/ICAC3N56670.2022.10074337.

[4] Q. A. Al-Haija and A. A. Badawi, "URL-based Phishing Websites Detection via Machine Learning," 2021 International Conference on Data Analytics for Business and Industry (ICDABI), Sakheer, Bahrain, 2021, pp. 644-649, doi: 10.1109/ICDABI53623.2021.9655851.

[5] M. Sameen, K. Han and S. O. Hwang, "PhishHaven—An Efficient Real-Time AI Phishing URLs Detection System," in IEEE Access, vol. 8, pp. 83425-83443, 2020, doi: 10.1109/ACCESS.2020.2991403

[6] A. Alswailem, B. Alabdullah, N. Alrumayh and A. Alsedrani, "Detecting Phishing Websites Using Machine Learning," 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2019, pp. 1-6, doi: 10.1109/CAIS.2019.8769571.

[7] S. Parekh, D. Parikh, S. Kotak and S. Sankhe, "A New Method for Detection of Phishing Websites: URL Detection," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 949-952, doi: 10.1109/ICICCT.2018.8473085.

[8] E. Buber, B. Dırı and O. K. Sahingoz, "Detecting phishing attacks from URL by using NLP techniques," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 2017, pp. 337-342, doi: 10.1109/UBMK.2017.8093406...

[9] Abdul Basit, Maham Zafar, Abdul Rehman Javed, Zunera Jalil, "A Novel Ensemble Machine Learning Method to Detect Phishing Attack", 2020 IEEE 23rd International Multitopic Conference (INMIC), pp.1-5, 2020.

[10] Hossein Shirazi, Shashika R. Muramudalige, Indrakshi Ray, Anura P. Jayasumana, "Improved Phishing Detection Algorithms using Adversarial Autoencoder Synthesized Data", 2020 IEEE 45th Conference on Local Computer Networks (LCN), pp.24-32, 2020.

[11] Hongpeng Zhu, "Online meta-learning firewall to prevent phishing attacks", Neural Computing and Applications, vol.32, no.23, pp.17137, 2020.

[12] Nguyet Quang Do, Ali Selamat, Ondrej Krejcar, Enrique Herrera-Viedma, Hamido Fujita, "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions", IEEE Access, vol.10, pp.36429-36463, 2020.

[13] Noor Faisal Abedin, Rosemary Bawm, Tawsif Sarwar, Mohammed Saifuddin, Mohammd Azizur Rahman, Sohrab Hossain, "Phishing Attack Detection using Machine Learning Classification Techniques", 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp.1125-1130, 2020.

[14] Jordan Stobbs, Biju Issac, Seibu Mary Jacob, "Phishing Web Page Detection Using Optimised Machine Learning", 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp.483-490, 2020.

[15] Suhas R. Sharma, Rahul Parthasarathy, Prasad B. Honnavalli, "A Feature Selection Comparative Study for Web Phishing Datasets", 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp.1-6, 2020.

[16] Krishna Mridha, Jahid Hasan, Saravanan D, Ankush Ghosh, "Phishing URL Classification Analysis Using ANN Algorithm", 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), pp.1-7, 2020.

[17] Rashmitha H.R., Sumana M., "Identification and Management of Frauds in Edge Computing Systems", 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp.1-5, 2020.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405124|

[18] Md. Sirajum Munir Prince, Asib Hasan, Faisal Muhammad Shah, "A New Ensemble Model for Phishing Detection Based on Hybrid Cumulative Feature Selection", 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp.7-12, 2020.

[19] J. Vijaya Chandra, Narasimham Challa, Sai Kiran Pasupuleti, "Detection of Deceptive Phishing Based on Machine Learning Techniques", Smart Technologies in Data Science and Communication, vol.105, pp.13, 2020.

[20] Smita Sindhu, Sunil Parameshwar Patil, Arya Sreevalsan, Faiz Rahman, Ms. Saritha A. N., "Phishing Detection using Random Forest, SVM and Neural Network with Backpropagation", 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), pp.391-394, 2020 Charu Singh, Meenu, "Phishing Website Detection Based on Machine Learning: A Survey", 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp.398-404, 2020.

[21] Hongpeng Zhu, "Online meta-learning firewall to prevent phishing attacks", Neural Computing and Applications, vol.32, no.23, pp.17137, 2020.

[22] Jordan Stobbs, Biju Issac, Seibu Mary Jacob, "Phishing Web Page Detection Using Optimised Machine Learning", 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp.483-490, 2020.

[23] Yang Chen, Rongfeng Zheng, Anmin Zhou, Shan Liao, Liang Liu, "Automatic Detection of Pornographic and Gambling Websites Based on Visual and Textual Content Using a Decision Mechanism", Sensors, vol.20, no.14, pp.3989, 2020.

[24] Ali Aljofey, Qingshan Jiang, Qiang Qu, Mingqing Huang, Jean-Pierre Niyigena, "An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL", Electronics, vol.9, no.9, pp.1514, 2020.

[25] Shaheen Mondal, Diksha Maheshwari, Nilima Pai, Ameyaa Biwalkar, "A Review on Detecting Phishing URLs using Clustering Algorithms", 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), pp.1-6, 2019.

[26] Mahajan Mayuri Vilas, Kakade Prachi Ghansham, Sawant Purva Jaypralash, Pawar Shila, "Detection of Phishing Website Using Machine Learning Approach", 2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), pp.384-389, 2019.

[27] Noor Syahirah Nordin, Mohd Arfian Ismail, Vitaliy Mezhuyev, Shahreen Kasim, Mohd Saberi Mohamad, Ashraf Osman Ibrahim, "Fuzzy Modelling using Firefly Algorithm for Phishing Detection", Advances in Science, Technology and Engineering Systems Journal, vol.4, no.6, pp.291, 2019.

[28] Y. Sönmez, T. Tuncer, H. Gökal and E. Avcı, "Phishing web sites features classification based on extreme learning machine," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, Turkey, 2018, pp. 1-5, doi: 10.1109/ISDFS.2018.8355342.

[29] L. Wu, X. Du and J. Wu, "Effective Defense Schemes for Phishing Attacks on Mobile Computing Platforms," in IEEE Transactions on Vehicular Technology, vol. 65, no. 8, pp. 6678-6691, Aug. 2016, doi: 10.1109/TVT.2015.2472993

S. Lee and J. Kim, "WarningBird: A Near Real-Time Detection System for Suspicious URLs in Twitter Stream," in IEEE Transactions on Dependable and Secure Computing, vol. 10, no. 3, pp. 183-195, May-June 2013, doi: 10.1109/TDSC.2013.3.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details