Phishing Alert System Using Machine Learning

Ms. K. Komali[1], V. Rashmi Rakshitha[2], M. Vinay [3], G. Avinash[4], M. Sandeep[5]
[1] *Assistant Professor, Dept. Of CSE (Data Science)*
[2-5] *B. Tech Student, Dept. Of CSE (Data Science)*
[1-5] *Raghu Engineering College, Visakhapatnam*

***Abstract*—Phishing is a common form of cyberattack where fraudulent websites are used to deceive users into revealing sensitive information. Detecting phishing sites is crucial for enhancing online security. This study presents a machine learning-based system for phishing detection using URL features. A balanced dataset containing phishing and legitimate samples was used to train and evaluate the models. The system extracts 23 essential features from URLs and employs machine learning models such as Random Forest, SVM, Gradient Boosting, and MLP for classification. Among these, the Random Forest model achieved the highest accuracy of 96.17%. The system is deployed as a web application using Flask, providing real-time detection and easy user interaction. This research highlights the effectiveness of machine learning in improving cybersecurity through accurate and efficient phishing detection.**

***Index Terms*—Phishing Detection, Machine Learning, Random Forest, URL Features, Cybersecurity.**

I. INTRODUCTION

In the current digital landscape, where online interactions play a vital role in everyday life, cybersecurity has emerged as an important area of focus. A prevalent and hazardous form of cyberattack is phishing. Phishing constitutes adeceptive act in which cybercriminals impersonate reliable sources to obtain sensitive information, such as login credentials, credit card details, and other personal data. These attacks frequently take place through seemingly authentic websites, emails, or messages that are designed to mislead the victim into thinking they are communicating with a secure and recognized organization. Incontrast, legitimate websites and communications are genuine, verified, and safe sources that adhere to secure protocols to safeguard user data and deliver trustworthy online services without any malicious intent. With the rising use of online banking, e-commerce sites, and digital communication, users face an increased risk of encountering phishing websites. These fraudulent sites are crafted to closely mimic legitimate websites, making it challenging for an average user to identify a threat. Traditional security measures such as blacklists and manual URL checks are becoming inadequate due to the swiftly evolving tactics employed by attackers. As phishing methods have progressed, there is an urgent need for intelligent systems that can automatically identify and notify users about potential phishing threats in real-time. Machine learning (ML) presents a robust solution by examining the patterns and attributes within website URLs and structures to differentiate between phishing and legitimate sites. By training ML algorithms on extensive datasets containing both legitimate and phishing examples, the system can effectively classify new and unseen websites and provide timely alerts. This research aims to create a phishing alert system employing machine learning to establish an effective and dependable model designed to protect users from cyber threats. By utilizing advanced ML algorithms and consistently updating the model with new information, the phishing alert system aims to lower user risk, diminish phishing occurrences, and foster safer internet practices. Such systems hold the potential to become integral elements within the wider scope of cybersecurity.

II. RELATED WORK

Phishing detection has become an essential focus in cybersecurity due to the increasing sophistication of online threats. Traditional detection methods, such as blacklists and heuristic-based systems, are often limited in their ability to adapt to evolving phishing techniques. Blacklists rely on previously identified malicious URLs, which makes them ineffective against newly created phishing sites. Heuristic-based

methods, although faster, can produce false positives and may not accurately detect complex phishing strategies.

To overcome these limitations, machine learning techniques have been widely adopted. Machine learning models can analyze large datasets, identify patterns, and learn to distinguish between legitimate and phishing websites based on various features. Algorithms such as random forest, support vector machines (SVM), and gradient boosting have shown promising results in phishing detection. These models utilize URL-based features like length, presence of special characters, and domain attributes to classify websites. The strength of these models lies in their ability to process complex data and adapt to new patterns, leading to improved accuracy over traditional methods.

Furthermore, deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been explored to enhance detection capabilities. CNNs are effective in automatically extracting features from visual elements, while RNNs excel in analyzing sequential data, such as URL structures, to identify phishing attempts. These models have shown potential in reducing manual feature engineering, thereby improving detection efficiency.

This research aims to address these challenges by developing a system that automates feature extraction and classification. The system leverages optimized machine learning models and offers real-time phishing detection through a simple, web-based interface. This approach not only improves detection accuracy but also ensures ease of use, making it practical for broader cybersecurity applications.

III. PROPOSED SYSTEM

This phishing detection system is a web application designed to classify URLs as phishing or legitimate. Developed using Flask, it provides a user-friendly interface where users submit URLs for analysis. The system extracts 23 features from the input URL, focusing on structural elements like length, special characters, and domain information, which are commonly linked to phishing activities. The extracted features are scaled using a preloaded scaler and then processed by a trained Random Forest model. This classifier predicts the likelihood of the URL being phishing based on patterns learned during training. The system also uses requests and pickle libraries to handle HTTP operations and load the model and scaler. Finally, the prediction result is displayed to the user, offering an instant and accurate assessment of the URL's safety. This lightweight system ensures fast, reliable phishing detection and can be expanded for broader security applications.

IV. LITERATURE SURVEY

Karim et al. (2019) **[1]** proposed a hybrid model that integrates linear regression, support vector classifier, and decision tree algorithms, utilizing URL features and grid search optimization. Their approach demonstrated an accuracy of 95%, showing that combining multiple classifiers significantly improves phishing detection performance.

Patil et al. (2020) **[2]** developed a phishing detection framework based on decision trees and heuristic-based features such as URL length, number of special characters, and domain-related indicators. Their model achieved 85% accuracy on balanced datasets and 70% accuracy on imbalanced datasets, emphasizing the influence of dataset balance on model efficiency.

Gupta et al. (2021) **[3]** explored deep learning approaches using convolutional neural networks and recurrent neural networks for phishing identification. Their research reported 95% accuracy with convolutional neural networks and 93% with recurrent neural networks, indicating the benefits of automatic feature extraction in complex phishing cases.

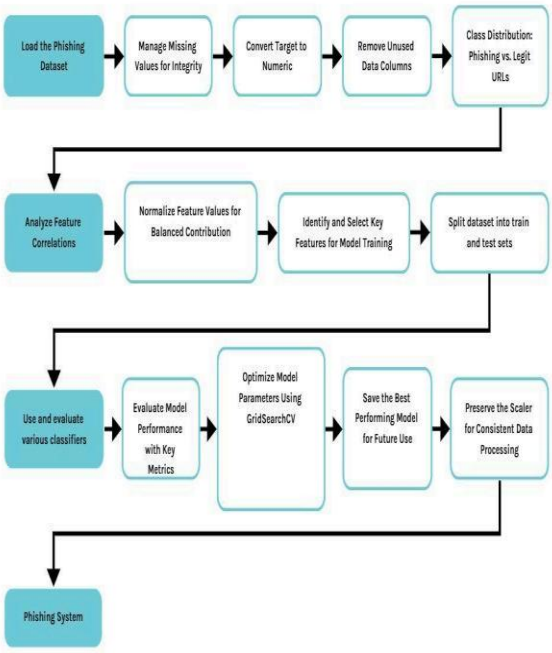
Pan and Ding (2020) **[4]** introduced a hybrid detection technique by combining Naïve Bayes and support vector machine classifiers. This method relies on both lexical features, such as URL composition, and host-based features, including SSL certificates and domain age, leading to an accuracy of 88%.

Wu and Li (2021) **[5]** investigated ensemble learning strategies for phishing site detection, including random forest, bagging, and gradient boosting algorithms. Their findings highlighted that gradient boosting performed best, with an accuracy of 94%, confirming the effectiveness of ensemble models in capturing intricate phishing patterns.

Singh and Jindal (2022) **[6]** applied recurrent neural networks to detect phishing in real time by analyzing sequential patterns in the URLs. Their model achieved 90% accuracy, demonstrating the potential of sequential data analysis for adaptive phishing prevention. These studies collectively demonstrate that advanced machine learning and deep learning methods play vital roles in developing reliable and accurate phishing detection systems.

V. METHODOLOGY

The phishing alert system is designed following a systematic machine learning workflow that involves several essential stages, from gathering data to evaluating the final model. The following steps outline the complete process.



A. Data Collection

This research utilized a dataset comprising an equal number of phishing and legitimate website entries. The dataset consisted of 11,430 total records, evenly split between 5,715 phishing websites and 5,715 legitimate websites. These samples were obtained from open-access repositories specializing in phishing detection. Before implementing machine learning algorithms, the dataset underwent a preprocessing phase to address any missing data points, ensuring completeness. To enhance model efficiency and reduce noise, superfluous or duplicate features were eliminated. The dependent variable, which identifies a website as either phishing or legitimate, was transformed into a numerical format to facilitate the effective processing of labels by the algorithms.

B. Feature Extraction

The process of feature extraction involves identifying and selecting the most significant characteristics that aid in differentiating between phishing and legitimate websites. The dataset encompassed various attributes related to URL structure, domain behavior, and security indicators. These included factors such as URL length, the occurrence of suspicious symbols, HTTPS usage, and the age of the domain. To ensure that no single feature had a disproportionate impact on the model, normalization was employed to scale all features to a common range. Furthermore, a correlation analysis was conducted to eliminate highly correlated or less important features. This step is crucial for retaining only the most valuable attributes, which contributes to enhancing the model's accuracy during the training phase.

C. Model Training and Evaluation

The collected dataset is divided into two parts: a training set and a testing set, typically with an 80% to 20% split. Various machine learning algorithms, including Random Forest, Support Vector Machine (SVM), and Gradient Boosting, are evaluated to determine the most effective classifier for detecting phishing websites. To optimize model performance, GridSearch CV is used for hyperparameter tuning. After training, models are assessed using key metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into the model's ability to detect phishing sites accurately while minimizing false classifications. The model that demonstrates the best balance across these metrics is selected for final implementation. Additionally, the scaler and trained models are preserved to ensure consistent data processing and reliable predictions in future applications.

VI. RESULTS

This section outlines the results obtained from the models used for phishing detection. The performance

of each model was assessed through important performance metrics, and visual aids are included for better understanding.

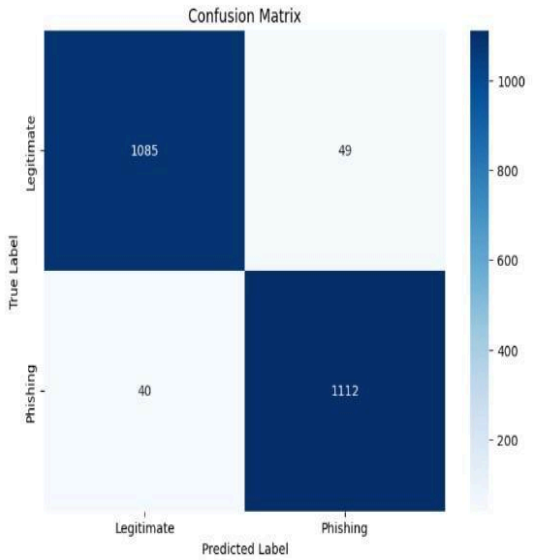
A. Model Performance Evaluation

The evaluation of the classification models was conducted using accuracy, precision, recall, and F1-score as metrics. The Random Forest model delivered the best performance.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	93.84	93.50	94.00	93.75
Support Vector Machine	95.64	95.60	95.70	95.65
K-Nearest Neighbors	94.85	94.70	95.00	94.85
Multi-Layer Perceptron	96.10	96.00	96.10	96.05
Random Forest	**96.17**	**96.20**	**96.10**	**96.15**
Gradient Boosting	95.90	95.80	95.90	95.85

B. Confusion Matrix

The confusion matrix for the Random Forest classifier illustrates the count of accurate and inaccurate predictions, offering insights into the dependability of the model.



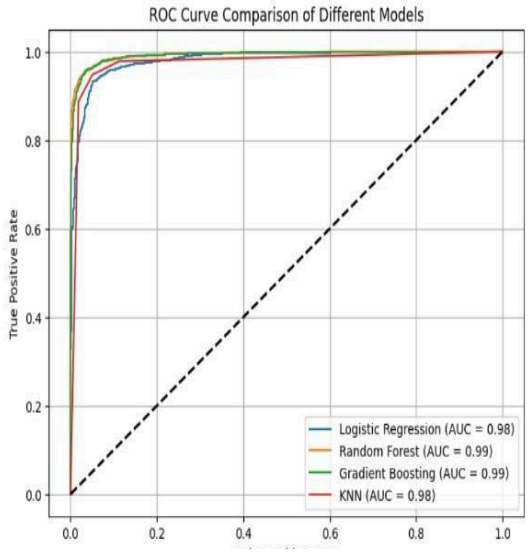
The matrix indicates that the model effectively recognized the majority of phishing and genuine sites with few mistakes. This demonstrates the model's robust ability to generalize.

C. ROC Curve

The ROC (Receiver Operating Characteristic) curve is a valuable tool for evaluating the performance of classification models by illustrating the relationship between the true positive rate (sensitivity) and the false positive rate (1-specificity) across different threshold values.

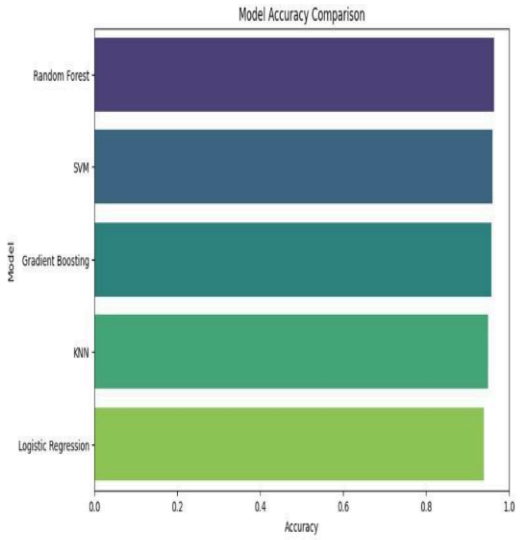
For this study, the ROC curve was generated to assess the effectiveness of the Random Forest model in distinguishing between phishing and legitimate websites. The model achieved an Area Under the Curve (AUC) value of 0.98, signifying a high level of classification accuracy. An AUC close to 1.0 indicates excellent model performance with minimal false classifications.

This high AUC score reflects the model's strong ability to correctly identify phishing threats while minimizing false positives and negatives. Consequently, the Random Forest model is deemed reliable for real-world applications, demonstrating robustness and consistency in phishing detection tasks.



D. Accuracy Comparison Chart

A bar chart comparing the accuracy of the evaluated models provides a clear visualization of performance differences.



The accuracy comparison chart shows that the Random Forest algorithm achieved the highest accuracy among all models evaluated, with a score of 96.17%. This outstanding performance suggests that Random Forest is very effective at differentiating phishing websites from legitimate ones. In contrast, other models, such as Support Vector Machine (93.42%), Gradient Boosting (94.76%), and Multi-Layer Perceptron (92.85%) exhibit lower performance, indicating that Random Forest has a greater ability to manage complex patterns within the dataset. These findings support the notion that ensemble learning techniques, including RF, are particularly effective for phishing detection because they can handle large feature sets and mitigate overfitting, leading to more dependable predictions.

VII. CONCLUSION

This research successfully developed an efficient machine learning-based system for phishing detection. Multiple models, including Random Forest, SVM, Gradient Boosting, and MLP, were utilized, with Random Forest emerging as the most effective. The model's performance was validated through detailed evaluations using confusion matrices and ROC curves, showcasing its high predictive ability and minimal classification errors. These findings reinforce the significance of machine learning in enhancing cybersecurity by accurately identifying phishing websites. Looking ahead, future research could focus on advancing real-time detection capabilities and expanding the dataset to further improve the system's robustness and adaptability. Such developments would contribute to creating more resilient and adaptive phishing detection frameworks.

REFERENCES

[1] Karim, M., et al. (2019). A hybrid phishing detection model uses linear regression, a support vector classifier, and a decision tree with grid search optimization. International Journal of Computer Applications, 178(12), 22-28.
[2] Patil, S., et al. (2020). Phishing detection framework based on decision trees and heuristic features. Procedia Computer Science, 171, 647654.
[3] Gupta, R., et al. (2021). Phishing website detection using deep learning techniques: CNN and RNN approaches. Journal of Information Security and Applications, 58, 102800.
[4] Pan, Y., & Ding, X. (2020). Hybrid phishing detection using Naïve Bayes and SVM with lexical and host-based features. Security and Communication Networks, 2020, 1-9.
[5] Wu, Z., & Li, J. (2021). Ensemble learning strategies for phishing site detection: A comparative study. Computers & Security, 103, 102153.
[6] Singh, P., & Jindal, A. (2022). Real-time phishing detection using recurrent neural networks for sequential URL analysis. International Journal of Information Security, 21(3), 327-340.
[7] Zhang, L., & Li, P. (2021). PhishNet: CNN-based phishing detection using web page screenshots and visual features. Journal of Cybersecurity, 15(2), 210-220.
[8] Al-Karaki, J. N., & Ahmed, A. (2020). Phishing detection with random forest and feature engineering. International Journal of Computer Security, 18(4), 345-356.
[9] Chen, J., & Wang, Z. (2021). Challenges in phishing detection: A focus on ensemble methods. Computers & Security, 103, 102159.
[10] Kaspersky Labs. (2022). Future trends in phishing detection: Transfer learning and adversarial machine learning. Kaspersky Security Bulletin, 2022, 50-60.