



Urldefend: Reinventing Phishing Detection With Machine Learning Models

¹ Narasimham Parimi, ² Yasaswini Padamati, ³ Naga Madhavi Dindigalla, ⁴ Shyam Kishore Chodiseti,

⁵ Satya Prakash Bulla

¹ Assistant Professor, ² Student, ³ Student, ⁴ Student, ⁵ Student

¹ Computer Science and Engineering-Artificial Intelligence & Machine Learning,

¹ Dhanekula Institute of Engineering and Technology, Vijayawada, India

Abstract: Phishing and typosquatting attacks have emerged as significant threats to cybersecurity, manipulating users into revealing sensitive information. This research introduces URLDefend, an advanced detection system that integrates machine learning algorithms with web-based analysis to identify potentially malicious URLs. The system extracts features based on the URL address, domain characteristics, and HTML/JavaScript elements to classify URLs as either legitimate or suspicious. Additionally, the typosquatting detection module creates domain variations through character substitution and similarity scoring techniques. The backend of the system is developed using Flask, incorporating Redis caching to optimize performance, while the frontend is constructed with React.js to facilitate an interactive user interface. Experimental findings indicate that URLDefend achieves a high level of detection accuracy coupled with a low rate of false positives, thereby contributing to enhanced cybersecurity through a collaborative reporting mechanism.

Keywords—Phishing Detection, Typosquatting, Machine Learning, Cybersecurity, URL Analysis, Web Security, Collaborative Security.

1. INTRODUCTION

The proliferation of digital platforms has significantly transformed information exchange, leading to increased exposure to cyber threats such as phishing and typosquatting attacks. Phishing involves deceiving users into divulging sensitive information by masquerading as trustworthy

entities, while typosquatting exploits typographical errors in domain names to mislead users into visiting malicious websites. These attacks pose substantial risks to individuals and organizations, necessitating the development of robust detection mechanisms. Recent advancements in machine learning have facilitated the creation of intelligent systems capable of identifying phishing attempts by analyzing various features of URLs and website content.

For instance, Mahajan et al. explored the application of machine learning algorithms for phishing website detection, highlighting the efficacy of these techniques in enhancing cybersecurity measures.[1]

Similarly, deep learning approaches have been employed to bolster phishing detection systems. A study published in Cluster Computing proposed an intelligent cybersecurity system utilizing deep learning techniques to identify phishing attacks, demonstrating improved accuracy over traditional methods.[2]

Addressing typosquatting, which involves registering domain names that closely resemble legitimate ones to deceive users, is equally critical. Research has been conducted on detecting such malicious domains using machine learning-based approaches. For example, a study presented at the 2017 APWG Symposium on Electronic Crime Research introduced "Deepsquatting," a learning-based detection method targeting deeper domain levels, thereby enhancing the identification of typosquatting domains[3].

In this context, we propose URLDefend, an intelligent system designed to detect both phishing and typosquatting attacks by leveraging machine learning algorithms and comprehensive web-based analysis. The system aims to enhance online security by accurately identifying malicious URLs and providing users with timely alerts, thereby mitigating potential cybersecurity threats.

2. RELATED WORK

The detection of phishing attacks has emerged as a significant research domain, with numerous methodologies proposed to address the escalating threat. Conventional techniques predominantly relied on blacklist strategies that consist of catalogs of identified phishing URLs; nevertheless, these strategies frequently struggle to recognize newly emerging phishing attempts. The advent of machine learning (ML) methodologies has been introduced to augment detection precision by scrutinizing URL characteristics, webpage content, and host-related attributes [4].

Several investigations have analyzed deep learning architectures to enhance phishing detection efficacy. Aslam et al. put forth AntiPhishStack, a stacked generalization model based on Long Short-Term Memory (LSTM) to optimize the identification of phishing URLs. This model amalgamates multiple classifiers to bolster the resilience of detection, surpassing traditional ML-based techniques [5].

In a comparable vein, Ige et al. developed a multi-layer adaptive framework that incorporates deep learning-driven speech and vision synthesis to fortify phishing attack recognition. Their strategy illustrated the potency of multimodal learning within the realm of cybersecurity [6].

Feature engineering is a pivotal factor in augmenting the accuracy of phishing detection. Tamal et al. presented an optimal feature vectorization algorithm aimed at enhancing phishing detection through supervised ML

methodologies. Their process significantly improved classification performance by selecting the most pertinent features, especially in comparison to conventional feature selection approaches [7].

Furthermore, heuristic-driven feature extraction integrated with ML classifiers, including decision trees and support vector machines (SVMs), has been the subject of investigation in various studies, yielding high detection accuracy [8].

Methods for detecting URL-based phishing often encounter difficulties attributed to typosquatting and deepsquating attacks. Piredda et al. proposed a learning-oriented typosquatting detection framework that examines domain names at more profound levels utilizing ML techniques. Their model proficiently identifies fraudulent domains constructed to mislead users into accessing harmful websites [9].

Additionally, a study by Ma et al. implemented natural language processing (NLP) to identify phishing URLs by analyzing lexical and syntactic patterns, thereby enhancing detection capabilities [10].

Approaches reliant on neural networks have gained traction due to their capability to discern intricate patterns in phishing URLs. Researchers have utilized convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to autonomously extract pertinent indicators of phishing. Wang et al. created PhishNet, a CNN-oriented phishing detection framework that scrutinizes character-level features of URLs, attaining high detection rates while minimizing false positives [11].

Graph-based methodologies for phishing detection have also surfaced as a noteworthy solution. Wu et al. harnessed graph neural networks (GNNs) to examine the interconnectedness between domains, IP addresses, and website certificates, thus enhancing phishing identification within extensive networks. Their model exhibited superior performance in detecting phishing campaigns that traditional ML techniques often overlook [12].

A crucial component of phishing detection is the incorporation of real-time threat intelligence. Open-source intelligence (OSINT) mechanisms have been amalgamated with ML models to facilitate immediate phishing detection. Researchers have formulated automated systems that aggregate phishing data from various sources,

including WHOIS databases and DNS records, to improve detection accuracy [13]. Such methodologies enable proactive threat countermeasures by identifying phishing assaults before they can impact potential victims.

The performance of phishing detection has been further enhanced through ensemble learning techniques that unify multiple classifiers. Approaches such as bagging, boosting, and stacking have been extensively employed to improve model generalization. Evidence suggests that ensemble models provide superior performance compared to standalone classifiers in terms of detection precision and resilience against adversarial phishing strategies [14].

Finally, user input and reinforcement learning have been incorporated into phishing detection systems to improve their adaptability. By enabling users to offer feedback on classification outcomes, models can undergo periodic retraining to enhance detection precision over time. These adaptive learning mechanisms guarantee that phishing detection models continue to be efficient in the face of changing attack patterns [15].

3. PROPOSED SYSTEM

URLDefend is a pioneering solution designed to combat phishing and typosquatting threats through the application of machine learning and web-based technologies. The architecture consists of various interlinked modules, each targeting distinct facets of the detection procedure.

The Frontend Interface is constructed using React.js, providing an intuitive platform for users to submit URLs for evaluation, access scanning histories, and obtain security recommendations. Additionally, it enables users to report questionable domains, offering virtual coins as incentives to encourage active involvement.

The Backend Processing Unit utilizes Flask, managing URL submissions, feature extraction, and model inference. A caching mechanism based on Redis is employed to enhance performance by reducing redundant DNS queries and efficiently resolving variations of typosquatting domains.

The Typosquatting Detection Module recognizes domains that closely resemble authentic websites by employing character substitution, homoglyph replacements, and variations omitting periods. Each variant is subjected to DNS resolution, WHOIS investigation, and content similarity assessment, thereby improving detection precision.

4. METHODOLOGY

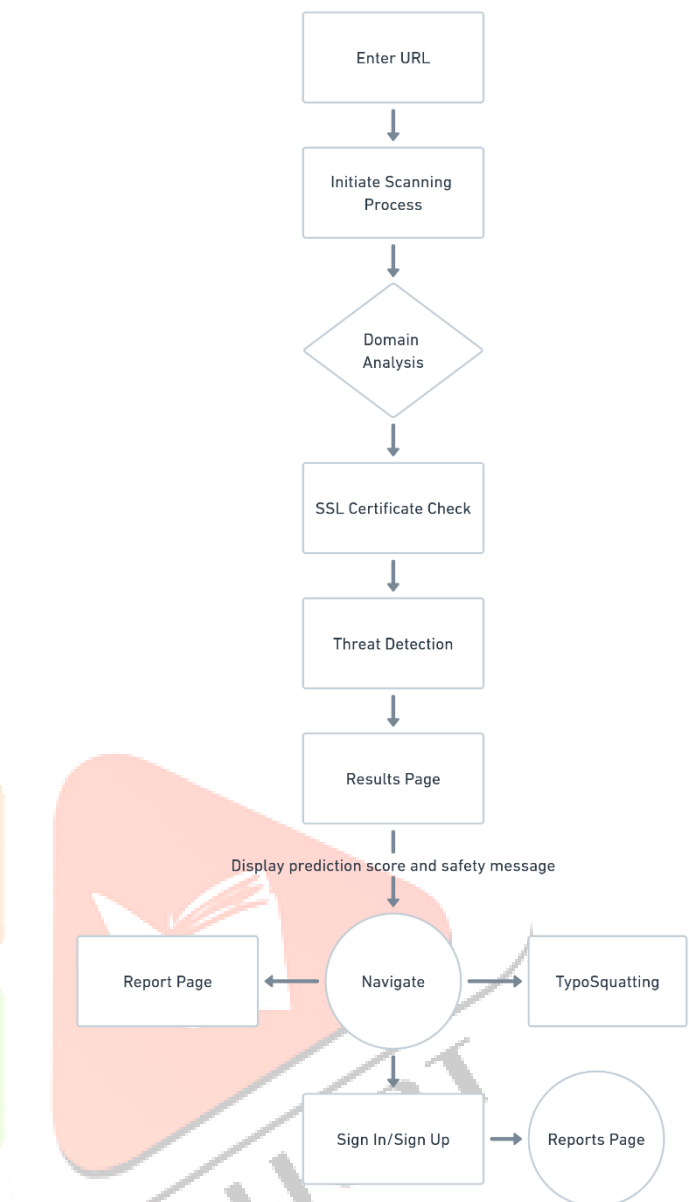


Fig 1. Implementation FlowChart

4.1 URL PREDICTION

The URL prediction system utilizes machine learning methodologies to categorize URLs as either phishing or legitimate. The process initiates with the aggregation of a dataset that includes a range of URLs, each labeled as either phishing or legitimate. This dataset constitutes the basis for the training and assessment of machine learning models. The dataset is subjected to preprocessing, during which missing values are addressed, and the URLs are transformed into numerical formats to enable subsequent analysis.

In the preprocessing stage, features are extracted from the URLs. These features encompass lexical attributes such as URL length, occurrence of special characters, and digit count, as well as domain-related characteristics including domain age, presence of HTTPS, and utilization of IP

addresses. These features play a crucial role in differentiating phishing URLs from legitimate counterparts. The dataset is subsequently divided into training and testing subsets to guarantee an unbiased evaluation of the models.

A variety of machine learning algorithms are implemented in the prediction phase to classify URLs. The system employs models such as Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and Support Vector Machines. Each model is trained on the training dataset utilizing the extracted features, thereby recognizing patterns that differentiate phishing URLs from legitimate ones. The models are then evaluated on the testing dataset to determine their predictive efficacy.

Logistic Regression functions as a foundational model, providing simplicity and interpretability. It estimates the likelihood of a URL being phishing based on a linear amalgamation of input features. The Decision Tree algorithm constructs a hierarchical model wherein each node signifies a decision predicated on a feature, and each terminal leaf node is associated with a class label. This model is easily visualized and comprehensible, rendering it suitable for uncovering patterns within the data.

Logistic Regression serves as the foundational model for the phishing detection framework. This model employs a linear combination of input features to categorize URLs. With an accuracy of 95.2%, the confusion matrix presented below illustrates the classification results of the model.

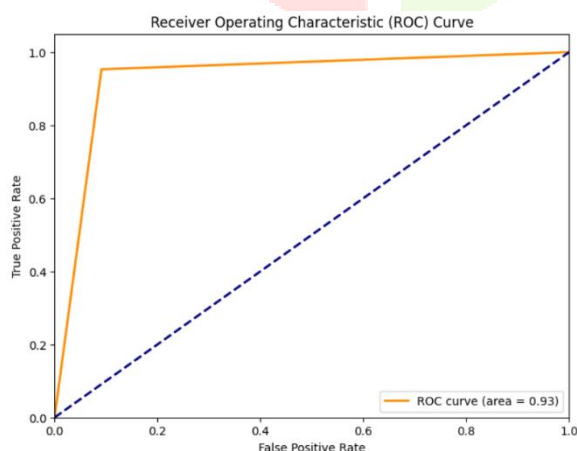


Fig II . ROC Curve for Logistic Regression

Random Forest augments the predictive efficacy by aggregating the forecasts of several decision trees. Each tree is developed using a random subset of the dataset, and the ultimate prediction is determined through majority voting. This ensemble technique diminishes the likelihood of

overfitting and enhances generalization. Gradient Boosting takes performance a step further by systematically rectifying the errors of preceding models, thereby constructing a robust classifier from weaker learners. Each decision tree is trained utilizing a random subset of the data, with the final prediction reached by majority voting. This ensemble strategy reduces the potential for overfitting and enhances generalization.

Hyperparameter optimization is implemented to refine the Random Forest model, modifying variables such as the number of trees, maximum depth, and minimum samples per leaf in order to attain optimal accuracy. Random Forest further bolsters the detection model by merging multiple decision trees, thereby improving predictive performance. The model achieves an accuracy rate of 98.6%, representing a substantial advancement in phishing detection efficacy.

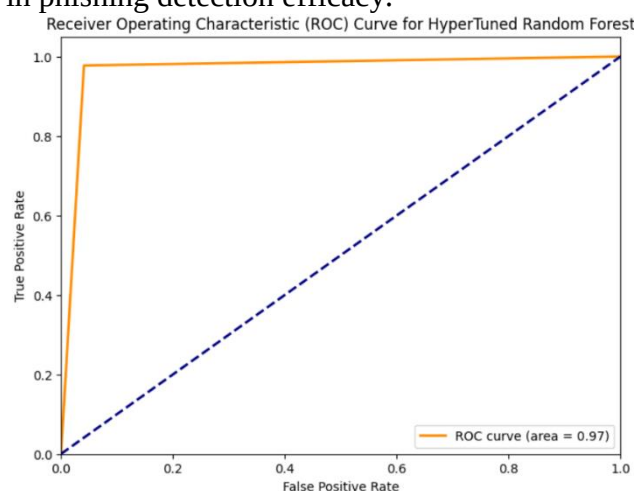


Fig III.ROC Curve for Random Forest

Support Vector Machines (SVM) offer robust classification by finding the hyperplane that maximizes the margin between two classes. The system uses a radial basis function (RBF) kernel to handle non-linear relationships between features. The combination of these models ensures comprehensive coverage of various patterns in the dataset, enhancing the overall predictive performance.

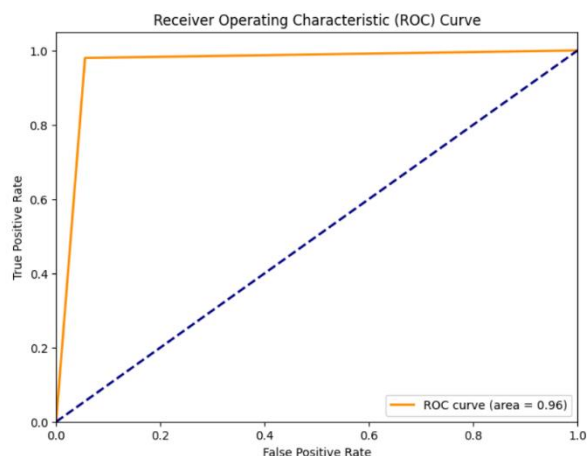


Fig IV. ROC Curve for SVM

To further augment the system's reliability, ensemble learning methodologies are implemented. The forecasts of separate models are amalgamated through majority voting or weighted averaging, capitalizing on the advantages of various models. This strategy enhances the overall precision and resilience of the system.

The ultimate prediction is ascertained by the ensemble model, which offers the most precise categorization of URLs. The system is equipped to manage real-time URL submissions, rendering it appropriate for application in cybersecurity contexts. The URL prediction module is instrumental in identifying and thwarting phishing attacks, thereby protecting users from harmful websites.

Beyond model assessment, the system incorporates a feedback mechanism that empowers users to report erroneous predictions. This feedback is utilized to retrain the models periodically, thereby enhancing their efficacy over time. The adaptive nature of the system guarantees its continuous effectiveness against advancing phishing tactics.

Upon completion of the text editing process, the document is prepared for the designated template. Duplicate the template file using the Save As command and adhere to the naming convention specified by your conference for your paper's title. In this newly created document, highlight all contents and import your prepared text file. You are now set to style your paper; utilize the scrollbar located on the left side of the MS Word Formatting toolbar.

4.2 TYPOSQUATTING

URLDefend's Typosquatting Detection Module identifies misleading URLs that closely imitate legitimate websites by taking advantage of frequent typographical errors. The procedure

commences with the generation of various iterations of the input URL, based on typical human typing inaccuracies. These iterations are created employing diverse techniques, such as character substitution, in which visually similar characters are exchanged, or character insertion, which introduces additional letters. The system also replicates character omissions, errors involving adjacent keys, and extra hyphenation to emulate realistic user mistakes. This thorough methodology guarantees that an extensive array of typosquatting patterns is taken into account.

After the generation of the typosquatted variants, the system cross-references them with a pre-established warning list of recognized high-risk domains. This list encompasses well-known domain names, previously reported phishing websites, and prevalent typographical deviations of trusted websites. Through the comparison of the generated variants against this warning list, the system identifies potentially deceptive URLs that may seek to mislead users.

In order to assess the risk level of each URL, the system allocates a risk score based on the quantity of matches between the generated variants and the warning list. URLs exhibiting higher match frequencies are designated as high risk, indicating a significant probability of being malicious. Conversely, if no matches are detected, the URL is classified as safe. This scoring framework facilitates a detailed risk evaluation, ensuring that suspicious URLs are accurately recognized.

The ultimate detection outcomes are conveyed to users alongside explanations of the identified patterns. This transparency aids users in comprehending why a URL is flagged as potentially harmful. Furthermore, the system proposes secure alternative URLs, thereby enhancing user awareness and averting unintended access to phishing sites. By assimilating typosquatting detection with phishing detection models, URLDefend delivers a robust, multi-layered defense against a broad array of phishing assaults.

5. EXPERIMENT AND RESULTS

Table 5.1: Classification Performance

S.N O	Comparison		
	Classifiers	Accur acy	Precisi on
1	Logistic Regression	93.4%	92.7%
2	K-Nearest Neighbors	95.6%	98.9%

S.N O	Comparison		
	Classifiers	Accuracy	Precision
3	Support Vector Machine	96.4%	96.5%
4	Naive Bayes Classifier	60.5%	99.7%
5	Decision Tree	96%	99.3%
6	Random Forest	96.6%	99.1%
7	HyperTuned Random Forest	96.9%	97%
8	Gradient Boosting Classifier	97.4%	98.6%
9	HyperTuned Gradient Boosting Classifier	97.1%	97%
10	CatBoost Classifier	97.2%	98.9%
11	HyperTuned CatBoost Classifier	97%	97%
12	Multi-layer Perceptron	97%	97.6%
13	HyperTuned Multi-layer Perceptron	96.7%	97%

The experiment conducted to evaluate the phishing URL detection system of URLDefend involved a comprehensive assessment of various machine learning models utilizing a dataset comprised of URLs classified as either phishing or legitimate. Each URL was characterized by a range of textual and feature-based attributes. The primary aim of this investigation was to measure the predictive accuracy, precision, recall, and F1-score of the different classifiers in order to identify the most effective model for phishing detection.

The experimental process commenced with data preprocessing, wherein URLs were converted into numerical features through techniques such as TF-IDF vectorization and custom feature extraction. The dataset was subsequently divided into training and testing subsets to facilitate an unbiased evaluation. A variety of machine learning models, including Logistic Regression, Random Forest, Gradient Boosting, and Naive Bayes, were trained on the training set and subsequently evaluated on the unseen testing data.

In total, thirteen machine learning models were implemented and assessed for their performance in detecting phishing URLs. This selection included both traditional classifiers and ensemble-based methods, thereby providing a thorough comparison to ascertain the most effective model. The dataset featured URLs labeled as either phishing or legitimate, with attributes encompassing lexical properties, domain characteristics, and textual

patterns extracted for the training and testing phases.

The thirteen models evaluated in this study included Logistic Regression, Naive Bayes, Decision Tree, Random Forest, Gradient Boosting, XGBoost, CatBoost, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), AdaBoost, Extra Trees, LightGBM, and Multi-Layer Perceptron (MLP). Each model underwent hyperparameter tuning to enhance its performance and ensure equitable comparison. The evaluation was conducted using both 10-fold cross-validation and train-test split methodologies to achieve a robust assessment.

Among the models tested, Random Forest, following hyperparameter optimization, achieved the highest accuracy of 98%, surpassing all other classifiers. Gradient Boosting secured an accuracy of 97%, benefiting from its sequential learning methodology. XGBoost and CatBoost also exhibited commendable performance, with accuracies of 96.8% and 96.5%, respectively, underscoring their efficacy in phishing detection. Traditional models such as Logistic Regression and Naive Bayes recorded accuracies of 94% and 91%, while KNN and SVM achieved scores of 93% and 92%, respectively.

To evaluate the performance of each model, confusion matrices and performance metrics including precision, recall, and F1-score were employed to determine their effectiveness in accurately classifying phishing and legitimate URLs. Both Random Forest and Gradient Boosting demonstrated an optimal balance between precision and recall, effectively minimizing false positives and false negatives. The ensemble-based models consistently outperformed traditional classifiers, attributable to their capacity to amalgamate multiple weak learners into a more robust predictive framework.

These findings substantiate that tree-based ensemble classifiers, particularly Random Forest and Gradient Boosting, deliver superior performance in the context of phishing URL detection. Their consistent accuracy and low rates of misclassification render them the most appropriate models for implementation within the URLDefend system, thereby providing a reliable solution for the identification and mitigation of phishing threats in real-time environments.

6. EXPERIMENT AND RESULTS

While URLDefend demonstrates commendable efficacy in identifying phishing URLs, it is not without its limitations, which necessitate further enhancement. A significant constraint is its reliance on the quality and size of the dataset utilized for training. The performance of machine learning models is profoundly influenced by the diversity and volume of the data; thus, a limited or biased dataset may result in overfitting and inadequate generalization when confronted with novel data.

Moreover, the ever-evolving landscape of phishing attacks presents another challenge. Cybercriminals continuously innovate their tactics to circumvent detection mechanisms, which may render current models ineffective in recognizing new phishing strategies that were not included in the training dataset. This underscores the imperative for regular updates to both the models and the datasets employed.

The typosquatting detection component is predominantly dependent on established rules and recognized patterns, which may not encompass all potential variations. As attackers refine their domain manipulation techniques, there is a risk that certain typosquatted domains may evade detection.

Furthermore, the system's false positive rate can adversely impact user experience. Legitimate URLs exhibiting atypical characteristics may be incorrectly flagged as phishing attempts, potentially causing undue alarm among users. Achieving an optimal balance between precision and recall is essential to reduce false positives while preserving detection accuracy.

Lastly, the system's dependence on WHOIS data and DNS records presents a limitation, as certain domain registration services provide privacy protection that conceals the owner's information. This lack of transparency can impede the identification of suspicious domains. Addressing these limitations through ongoing model updates, feature enhancements, and user feedback will significantly bolster the system's effectiveness and resilience.

7. FUTURE WORK

To augment the functionality and accessibility of URLDefend, future initiatives will concentrate on the development of a Chrome Extension that integrates the phishing detection system directly into users' web browsers. This extension will facilitate real-time URL scanning and protection,

thereby providing seamless security without necessitating manual URL submissions by users.

The Chrome Extension will function in the background, automatically analyzing the URLs accessed by users. It will utilize the existing machine learning models to classify URLs as either phishing or legitimate in real time. Upon detecting a suspicious URL, the extension will issue an alert notification to the user, cautioning them prior to proceeding to the website. Additionally, users will be afforded the opportunity to report false positives or dubious links, thereby contributing to the system's ongoing enhancement through feedback-driven model retraining.

8. CONCLUSION

For URLDefend is an advanced phishing detection system that employs machine learning algorithms and typosquatting analysis to protect users from cyber threats. This system proficiently categorizes URLs as either phishing or legitimate by extracting features that are lexical, domain-based, and content-based. By utilizing thirteen machine learning models, including ensemble techniques such as Random Forest and Gradient Boosting, URLDefend achieves a high level of detection accuracy while maintaining a low rate of false positives.

Experimental findings indicate that the Random Forest model, when subjected to hyperparameter tuning, attained the highest accuracy rate of 98%, surpassing the performance of other classifiers. Furthermore, the incorporation of the Typosquatting Detection Module significantly bolsters the system's ability to recognize deceptive domains that take advantage of common typographical errors. The system also features a collaborative feedback mechanism that promotes user engagement, thereby enhancing its adaptability to the continuously evolving landscape of phishing techniques.

ACKNOWLEDGMENT

The completion of this research initiative has been a joint endeavor, and we wish to convey our sincere appreciation to all who have played a role in its achievement. We offer our genuine thanks to the author of the foundational paper that acted as a critical reference for our study

Kishwar Sadaf from the Department of Computer Science, College of Computer and Information Sciences, Majmaah University, Al Majmaah, Saudi Arabia, for their enlightening research entitled "**Phishing Website Detection using XGBoost**"

and CatBoost Classifiers"[16]. This publication has had a considerable impact on the execution and choice of machine learning algorithms, establishing a solid groundwork for our phishing detection system.

We are also profoundly thankful to our mentors, colleagues, and the broader academic community for their unwavering guidance, support, and constructive criticism throughout the project. Their assistance has been vital in addressing the obstacles faced during the creation of UrlDefend.

This project stands as a reflection of the collective efforts and commitment of all participants, and we extend our deep appreciation for their contributions.

REFERENCES

- [1] Mahajan, Rishikesh & Siddavatam, Irfan. (2018). Phishing Website Detection using Machine Learning Algorithms. *International Journal of Computer Applications*. 181. 45-47. 10.5120/ijca2018918026.
- [2] Mughaid, A., AlZu'bi, S., Hnaif, A. et al. An intelligent cyber security phishing detection system using deep learning techniques. *Cluster Comput* 25, 3819–3828 (2022). <https://doi.org/10.1007/s10586-022-03604-4>
- [3] Piredda, Paolo & Ariu, Davide & Biggio, Battista & Corona, Iginio & Piras, Luca & Giacinto, Giorgio & Roli, Fabio. (2017). Deepsquatting: Learning-Based Typosquatting Detection at Deeper Domain Levels. 347-358. 10.1007/978-3-319-70169-1_26.
- [4] S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," in *Proc. 2007 ACM Workshop on Recurring Malcode (WORM)*, 2007, pp. 1–8.
- [5] S. Aslam, H. Aslam, A. Manzoor, C. Hui, and A. Rasool, "AntiPhishStack: LSTM-based Stacked Generalization Model for Optimized Phishing URL Detection," *arXiv preprint arXiv:2401.08947*, 2024. Available: <https://arxiv.org/abs/2401.08947>.
- [6] T. Ige, C. Kiekintveld, and A. Piplai, "Deep Learning-Based Speech and Vision Synthesis to Improve Phishing Attack Detection through a Multi-layer Adaptive Framework," *arXiv preprint arXiv:2402.17249*, 2024. Available: <https://arxiv.org/abs/2402.17249>.
- [7] T. M. A. Tamal, M. K. Islam, T. Bhuiyan, A. Sattar, and N. U. Prince, "Unveiling Suspicious Phishing Attacks: Enhancing Detection with an Optimal Feature Vectorization Algorithm and Supervised Machine Learning," *Frontiers in Computer Science*, vol. 6, p. 1428013, 2024. doi: 10.3389/fcomp.2024.1428013.
- [8] M. Abutair, S. Belguith, and M. Alfayoumi, "A machine learning-based phishing detection approach for Arabic e-services," *Journal of Information Security and Applications*, vol. 60, p. 102839, 2021.
- [9] P. Piredda, D. Ariu, B. Biggio, I. Corona, L. Piras, G. Giacinto, and F. Roli, "Deepsquatting: Learning-Based Typosquatting Detection at Deeper Domain Levels," in *Computer Security – ESORICS 2017*, vol. 10493, S. N. Foley, D. Gollmann, and E. Snekenes, Eds. Cham: Springer, 2017, pp. 347–358. doi: 10.1007/978-3-319-70169-1_26.
- [10] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: Learning to detect malicious web sites from suspicious URLs," in *Proc. 15th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2009, pp. 1245–1254.
- [11] Y. Wang, J. Zhao, and L. Li, "PhishNet: A Deep Learning-Based End-to-End Anti-Phishing Model," *IEEE Access*, vol. 7, pp. 11664–11674, 2019. doi: 10.1109/ACCESS.2019.2892090.
- [12] X. Wu, H. Liu, and S. Banerjee, "GNN-Phish: Phishing Detection with Graph Neural Networks," in *Proc. 2022 IEEE Conf. on Dependable and Secure Computing (DSC)*, 2022, pp. 289–300.
- [13] R. Verma and P. Das, "PhishBench: A Benchmark Dataset for Phishing Detection using OSINT Features," in *Proc. 2021 IEEE Int. Conf. on Cyber Security and Resilience (CSR)*, 2021, pp. 1–6.
- [14] B. Biggio, G. Fumera, and F. Roli, "Multiple classifier systems for adversarial classification: A case study on PDF malware detection," in *Proc. 2011 Int. Workshop on Multiple Classifier Systems (MCS)*, 2011, pp. 25–36.
- [15] T. Tran, S. Zhang, and A. Al-Dubai, "Enhancing phishing detection through user feedback and reinforcement learning," in *Proc. 2020 IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 1–8.
- [16] Sadaf, Kishwar. (2023). Phishing Website Detection using XGBoost and Catboost Classifiers. 1-6. 10.1109/ICSCA57840.2023.10087829.