Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 28 No. 4s (2025)

An Ensemble Approach to Phishing URL Detection Using Supervised Machine Learning

Dr **Eldho K J**[1], Mr **V Surendhiran**[2], **Dr.S. Karuppusamy**[3], **Dr. P. Vijayakumar**[4]

[1]Assistant Professor, Department of Computer Science, Mary Matha Arts & Science College Wayanad, India.
eldhokj@marymathacollege.ac.in

[2]Assistant Professor, Department of Computer Science and Engineering, Excel Engineering College, Namakkal, India.
surendhiran.svs@gmail.com

[3]Associate Professor, Department of Computer Science and Engineering, Nandha Engineering College, Erode, India.
sksamymsc@gmail.com

[4]Assistant Professor, Department of Computer Applications, Kongu Engineering College, Erode, India.
vijayakumarpalanisamy@gmail.com

Article History:

Received: *14-10-2024*

Revised: *21-11-2024*

Accepted: *02-12-2024*

Abstract:

Phishing is a cybercriminal activity which deceives users into providing their private information, such as credit card credentials and sensitive information like passwords which ultimately leads to financial loss. Phishing attacks are still a serious issue in the digital world because they try to trick people into divulging important information. This study aims to address this difficulty by employing supervised machine learning to detect phishing URLs.To improve accuracy, classification methods are specifically utilized. To train and evaluate the classifiers, the study methodology makes use of a dataset that includes benign and phishing URLs along with a variety of attributes. The study compares the effectiveness of conventional single classifiers with ensemble classifiers, such as Random Forest, Gradient Boosting, and CatBoost. The experimental results show that the ensemble classifier performed better than the individual classifiers and detected phishing URLs more effectively and with a higher accuracy. The suggested method emphasizes the value of using many classifiers to increase detection accuracy and robustness, showcasing the efficiency of ensemble learning techniques in improving the classification problem. This research aims to advance cyber security measures by proving the effectiveness of ensemble classifiers in phishing URL detection and provides a dependable and effective way to counteract evolving phishing threats.

Index Terms—Phishing, URL, Machine learning, Classifers, Ensemble model, Supervised Learning

1. Introduction

The threat of cyber attacks is increasing exponentially in today's digital environment, with one of the primary threat vectors being the rise of phishing. As the weakest link in the cyber security chain, human users are the target of phishing. Attackers try to trick people into disclosing private information, committing identity theft, financial fraud, or ob taining unauthorized access to systems by taking

advantage of human weaknesses like trust and curiosity. Traditional phishing techniques frequently depended on trick emails or messages, but as technology has advanced, so has the number of phishing URLs. Malicious web links intended to look like trustworthy websites trick unwary users into sending sensi tive information. Given the frequency and sophistication of phishing attacks, cyber security experts face a huge problem in detecting these malicious URLs. Researchers and practitioners have been using machine learning techniques more frequently as a protection against phishing attacks in response to this expanding threat.

Machine learning presents the possibility of automated, scalable solutions for spotting phishing URLs with more accuracy and efficiency than conventional methods by utilizing large datasets and sophisticated algorithms. This work explores the identification of phishing URLs along with an emphasis on supervised machine learning techniques. Our objective is to differentiate between secure and harmful URLs using classification methods, enabling the early identification and prevention of attacks that involve phishing. In addition, we assess how well ensemble classifiers perform in improving detection accuracy, providing insight into the effectiveness of integrating several algorithms to counteract sophisticated phishing threats. The application of machine learning tech niques were selected due to their superiority over conven tional methods like blacklisting/whitelisting and heuristics approach. Unlike heuristic-based approaches, which typically rely on rule-based systems and struggle to adapt to evolving phishing techniques,machine learning provides a flexible and dynamic approach that can identify patterns and trends in data.

The recognition of phishing URL detection as a classification task is fundamental to our methodology. Making use of this knowledge, we carefully select datasets that have a fair proportion of phishing and trustworthy URLs. The chosen classifiers will be trained on a representative and diversified sample of balanced dataset, which will improve their capacity to distinguish between benign and malicious URLs. Classifiers can identify tiny differences that can indicate malicious intent and capture nuanced patterns owing to the extensive feature collection. We give algorithms the capability to distinguish be tween phishing and authentic URLs by integrating an extensive range of variables. Classifiers are trained using a supervised machine learning approach in our methodology. We assess the effectiveness of our classifiers by assessing their performance using accepted metrics like accuracy value, precision, recall, and F1-score. Future upgrades and modifications will be driven by the analysis's insightful findings about the detection system's benefits and drawbacks.

2. Literature review

The literature survey is based on papers that has been selected on the basis of their relevance and innovative ap proach that has been proposed. Although various traditional methods such as blacklisting, heuristics based approach are still in existence by the commonly used web browsers, the rising statistics of financial frauds and data loss points to the immediate need of flexible,spontaneous and effective methods of detection. A solution based on machine learning, a cutting edge subset of artificial intelligence is proposed to reduce the ever rising problem of phishing attacks and threats. Several investigations have previously been conducted within this field in order to address the matter. Currently new ways to tackle the issue is being discussed and studied worldwide to combat the issue.

An exponential surge in cybercrimes has been documented in the wake of the internet revolution, with phishing attacks ranking among the most widespread.SMS/text phishing, voice phishing, spam phishing, and website phishing, among others, have become common among us. Karim et al. [1] proposes a hybrid machine learning approach utilizing various models to identify and prevent phishing URLs with remarkable precision and efficiency. The paper addresses multiple machine learning models, which include support vector classifier, naive Bayes, gradient boosting classifier, K-neighbors classifier, decision tree, linear regression, random forest, and a hybrid LSD model-a combination of logistic regression, support vector machine, and decision tree. The proposed strategy involves a hybrid LSD in conjunction with the canopy feature selection technique, cross-fold validation, and Grid Search Hyperpa rameter Optimization approaches. The results show that, with a high degree of accuracy and efficiency, the recommended approach surpasses other models and produces the most favor able outcomes in providing a shield against phishing attacks. To filter phishing URLs, Indrasiri et al.[2] proposed a robust ensemble machine learning model called the Expandable Ran dom Gradient Stacked Voting Classifier (ERG-SVC), which uses two ensemble techniques, one clustering algorithm, seven classification algorithms, and supervised and unsupervised techniques. The proposed model integrates predictions from multiple ML models through a stacking classifier.

In the paper proposed by Sameen et al.[3] PhishHaven, a detection system that distinguishes between artificial intelli gence (AI)-generated and human-crafted phishing URLs using a collection of machine learning algorithms are presented.In order to improve analysis, the system integrates URL HTML Encoding, a URL Hit method, and lexical analysis for feature extraction. Artificial Intelligence (AI) techniques, namely the meta-learner models, are proposed by Alsariera, Adeyemo et al. [4]. These models are constructed utilizing the extra-tree base classifier. The methodology involves the development of the models,including the selection of appropriate meta-learner algorithms and base learners. The extra-tree base classifier is used in the development of metalearner models BET, ABET, RoFBET, and LBET to identify phishing websites.This dataset of phishing websites was used to fit the suggested AI-based meta learner models and assess how well they identified phishing activity.The models exceeded other machine learning based models in terms of detection accuracy, with a low false positive rate of 0.028 and high detection accuracy. PDGAN, a phishing detection technique proposed by Al-Ahmadi et al. [5], uses a website's URL to provide dependable performance. Adiscriminator network and a generator network make up this paradigm. The generator network generates synthetic phishing URLs, while the discriminator network assesses if the URLs are phishing or legitimate. CNN and LSTM are combined in the PDGAN model. The model uses a pre-designed model to classify a URL after encoding it into a two-dimensional tensor.By examining how character-level similarity affects the creation of artificial phishing URLs, the approach seeks to improve the categorization of phishing URLs.It also discusses the several approaches such as machine learning, deep learn ing, heuristic-based, and visual similarity based that are used to identify phishing URLs. Using a machine learning method, classification algorithms are trained and feature representations are obtained from URLs. Creating a deep learning model and choosing model inputs are the main goals of the deep learning methodology.

Alshingiti et al.[6] presented a system that consists of four phases: pre-processing, training three different models (LSTM, CNN, and LSTM-CNN), extracting features from URLs, and f inally classifying webpage URLs as phishing or real. The sys tem exhibits its resilience by identifying

phishing URLs using two different methods. For comparison, the two deep learning models—LSTM and CNN are used independently, and an LSTM-CNN based approach is also advised. The experimental f indings shows the effectiveness of CNN and LSTM-CNN both achieving the highest levels of accuracy. The research authored by Aldakheel et al. [7] suggest a unique method based on a Convolutional Neural Network (CNN) model for accurately identifying phishing attempts.This methodology works quite well at differentiating between authentic and fraudulent web sites. The PhishTank dataset, well known dataset for URL based p detection is used by the authors to evaluate their methodology. The methodology consists of four discrete steps. The preparation and preprocessing of the dataset is the main focus of the first stage. The subsequent phase is distinguished by the transformation of URL data into a character vector by use of character embedding. Using a Convolutional Neural Network (CNN) to examine the character vectors is the third phase. The recovery of the URL attributes, which are essential for understanding the inner workings of the trained model, is the focus of the fourth and final stage. Their method outperforms earlier state-of-the-art models with high accuracy rates.

The paper proposed by Yang et al.[8]puts forward a method for identifying phishing websites. Existing techniques for anti-phishing have limitations in terms of feature extraction expertise and time, as well as delays caused by third party services. For the purpose of phishing website identification, the suggested technique combines random forest (RF) with convolutional neural networks (CNN) [8]. Fixed-size matri ces are created by converting URLs using character embed ding techniques. After CNN models are employed to extract features, multiple RF classifiers are utilized to classify the characteristics in order to forecast the validity of URLs. The study includes information on the datasets utilized and perfor mance indicators measured in the tests carried out to assess the efficacy of the suggested approach. Strong generalization ability, independence from third-party services, and language independence are the benefits of the suggested approach. The suggested method's usage of CNN, RF, and character embedding to identify phishing websites is one of its key features. The study also emphasizes how well the suggested strategy performs in terms of accuracy and false positive rate, as well as its advantages, such as strong ability for general ization, independence from third-party services,and language independence. The study put forth by Adebowale et al.[9] has focused on the development and conceptualization of a deep learning-based phishing detection solution, utilizing website and url content, which includes text, graphics, and frames.The paper presents an investigation centered on the fusion of hybrid characteristics derived from textual, pictorial, and sequential data to construct a resilient deep learning mechanism for the purpose of detecting phishing [9].It reveals the significance of assimilating image,text, and frame attributes in phishing detection systems to enhance their efficiency.The research pro poses the adoption of deep learning algorithms, LSTM+CNN, to produce an amalgamated solution for phishing detection.The study involves the hybrid approach of employing both CNN and LSTM algorithms alongside the combination of image, frame, and text features.

The study presented \by Kumar et al. [10] focuses on the increasing problem of phishing websites, which poses an important challenge for internet service providers. The Swarm Intelligence Binary Bat Algorithm is a novel approach that the authors provide in their paper with the goal of building a neural network that can categorize network URLs, with a focus on detecting phishing websites. The methodology comprises vari ous steps for detecting and classifying phishing websites. The process

initiates with the acquisition of a benchmark dataset, which entails samples from both legitimate and phishing web sites. The data undergoes preprocessing to eliminate features with missing or null values. Relevant attributes associated with phishing URLs are then extracted. The proposed SI-BBA al gorithm, utilizing swarm intelligence, is applied for parameter extraction and classification. The methodology encompasses data acquisition, preprocessing, feature extraction, algorithm implementation, and results evaluation. The empirical findings shows the effec- tiveness of this approach,with an Adam opti mizer attainingthe Highest classification accuracy in detecting phishing website attacks by utilizing swarm intelligence techniques.

3. Proposed approach

After an evaluation of various studies conducted in the domain of phishing detection, this work proposes a system utilising the effective usage of machine learning.Various ma chine learning algorithms have been proven to successfully tackle the issue of phishing.In this work several machine learning algorithms are tested against the chosen dataset and are evaluated.Based on the results obtained we find an effective tool to overcome the threat of phishing attacks.

A. System Design

The system design comprises of several key phases ranging from selection of appropriate dataset to obtaining the final prediction. It initiates with data cleaning and preprocessing to ensure data quality. Subsequently, model selection is crucial for effective phishing detection. A process flow diagram for the system is depicted in the figure1

Fig. 1. System Design - Phases

B. Data Selection

For our study on phishing URL identification, we col lected two substantial data sets from reputable public sources to aid in the achievement of our research objectives. Our study's foundation is mostly formed by the first dataset [11], which we got from Kaggle and has 11,055 instances with 30 attributes. These features cover a wide range of URL attributes, enabling a thorough understanding of phishing URL patterns.11,430 instances and 86 characteristics make up the bigger feature set of the second dataset [12], which was obtained via Mendeley. This dataset offers a more in-depth analysis of the structural and semantic components of URLs, which increases the accuracy of our research. This dataset contributes to the development of robust detection models and enhances our understanding of phishing URL behaviors by providing a greater number of variables, such as detailed domain information.

Our priority was to ensure data quality, reliability, and relevance to our study goals when gathering these datasets. Our objective is to create sophisticated machine learning models for phishing URL identification by utilizing these carefully selected datasets. Diverse dataset properties, such as different feature sets and instance sizes, make it easier to thoroughly test and validate detection techniques. We

lay a strong foundation for thorough and perceptive research in the field of cybersecurity and phishing detection with our systematic approach to dataset preparation and gathering.

C. Data Cleaning

Preparing the gathered dataset for analysis is the goal of the data cleaning process. This include dealing with any missing values, eliminating duplicates, and fixing any dataset discrep ancies. It's critical to find and fix problems with phishing URL data, such as abnormalities in domain information, inconsistent labeling, or irregularities in URL structures. The remaining phases are influenced by the dataset's integrity.

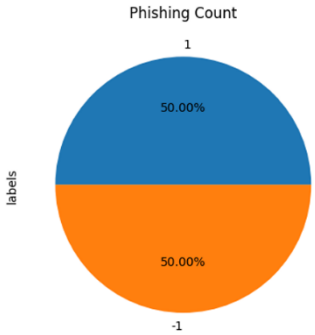

Fig. 2. Distribution of URLs-Dataset-1

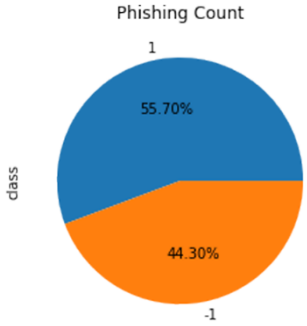

Fig. 3. Distribution of URLs-Dataset-2

D. Data Pre-Processing

Data preprocessing is a crucial step where the prepared dataset undergoes transformations to make it suitable for machine learning algorithms. For phishing URL data, this may involve feature scaling, encoding categorical variables, and extracting relevant information from URLs. Normalizing and standardizing features ensure that the machine learning model can effectively interpret and learn from the data, contributing to improved model performance.

E. Model Selection

Machine learning model selection is a pivotal step in design ing an effective and accurate system for phishing detection.The fundamental purpose of fraudulent URL detection is to clas sify URLs according to different attributes into legitimate or phishing classes. Selecting a suitable algorithm,or group of algorithms,that can efficiently extract patterns from the data and generate precise predictions, is known as model selection.Several machine learning algorithms are particularly suited to classification problems like phishing URL detection. The following is a list of such algorithms:

• Gradient Boosting Classifier

• CatBoost Classifier

• Multi-layer Perceptron

• XGBoost Classifier

• Random Forest

• Support Vector Machine

• Decision Tree

• K-Nearest Neighbors

• Logistic Regression

• Naive Bayes Classifier

Models were trained on each of these classifiers on the chosen datasets.From experimental results it was evident that the Gradient Booster Classifier Model attained the highest level of accuracy.

F. Ensemble Approach in Machine Learning

Ensemble models are a powerful technique in machine learning where multiple models are combined to make pre dictions. The idea behind ensemble learning technique is to achieve better performance than any single model on its own. This is because ensemble models can compensate for the weaknesses of individual models and leverage their strengths. Their strengths include:

• Reduced Variance: Ensemble models decrease the vari ance of predictions, particularly beneficial when individ ual models tend to overfit.

• Improved Robustness: By combining multiple models trained on diverse subsets of data or employing different algorithms, ensemble models exhibit greater resistance to noise and outliers.

• Enhanced Performance: Ensemble models frequently out perform individual models in predictive accuracy, partic ularly evident in intricate and varied datasets.

There are various kinds of ensemble architectures:

• Bagging (Bootstrap Aggregating): It involves training multiple instances of a base model on different subsets of the training data, typically sampled with replacement. The final prediction is usually the average (for regression) or majority vote (for classification) of predictions from individual models.

• Boosting: Boosting algorithms sequentially train models, each focusing on instances that were misclassified by previous models. Popular boosting algorithms include AdaBoost, Gradient Boosting Machines (GBM), XG Boost, and LightGBM.

• Stacking (Stacked Generalization): Stacking combines predictions from multiple base models by training a meta model on their outputs. Base models are trained on the original dataset, and the meta-model learns to combine their predictions optimally.

• Voting: In a voting ensemble, multiple models make predictions, and the final prediction is determined by majority voting for classification, or averaging in case of

• Blending: Similar to stacking, blending combines predic tions from multiple models, but instead of using a meta model, it uses a simple algorithms like linear regression to combine their outputs.

We utilise ensemble models in this project as ensemble models can greatly enhance phishing URL detection by combining the predictions of multiple base learners, each trained on different aspects of the data. By utilising diverse algorithms such as de cision trees, random forests, support vector machines, and neu ral networks, ensemble methods can effectively capture various characteristics of phishing URLs. For instance, decision trees excel at identifying specific patterns in URL structures,

while neural networks might uncover subtle correlations in feature representations. Combining these diverse models can mitigate individual model biases and improve overall performance. Additionally, ensemble methods provide robustness against noisy or ambiguous data, making them well-suited for the dynamic and evolving nature of phishing attacks. We utilise the machine learning libraries such as scikit-learn for the built-in support for ensemble models.

G. Model Training

The dataset was used for iterative training of each clas sifier, with hyperparameter optimization used to maximize performance.The model demonstrating the highest accuracy and robust performance across multiple metrics was selected as the primary candidate.

4. Results and discussion

The model selection was done based on the evaluation of various algorithms trained on the chosen datasets.The results of evaluation are depicted in the fig4 and fig5

The tableI below shows the perfomance metrics values of various models on kaggle dataset.

Table 1: results of various models

Models	Accuracy (%)	F1 (%)	Recall (%)	Precision (%)
Gradient Booster	97.4	97.7	99.4	98.6
Random Forest	96.7	97.0	99.2	98.5
Decision Tree	96.1	96.5	99.1	99.3
KNN	95.6	96.1	99.1	98.9
Logistic Regression	93.4	94.1	94.3	92.7
XG Boost	96.9	97.3	99.3	98.4
CatBoost	97.4	97.7	99.4	98.6
SVM	96.4	96.8	98.0	96.5

Based on experimental results it was evident that Gradient Booster classifier have a higher rate of efficiency.The accuracy of 97.4% and an F1 score of 97.7% ensures reliable perfor mance in handling false positives and false negatives. Its recall rate of 99.4% proves its ability to effectively identify a signif icant portion of actual phishing URLs, minimizing the risk of overlooking potential threats. The precise predictions, reflected in a precision of 98.6%, affirm the model’s accuracy in labeling positive instances. Moreover, as a boosting algorithm, Gradient Boosting leverages ensemble learning, combining the strengths of multiple weak learners for enhanced predictive performance and adaptability to various data patterns. Its robustness in han dling complex relationships within the data further establishes its suitability for the desired task of phishing URL detection. The figure6 visualises the higher values of evaluation metrics for the datasets considered.

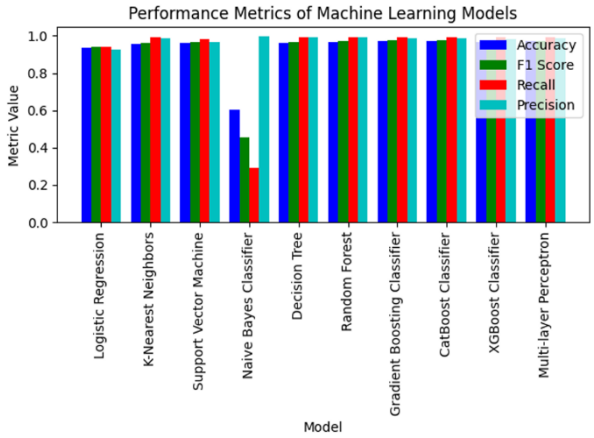

Fig. 4. Perfomance Metrics of Machine Learning Models-Dataset-1

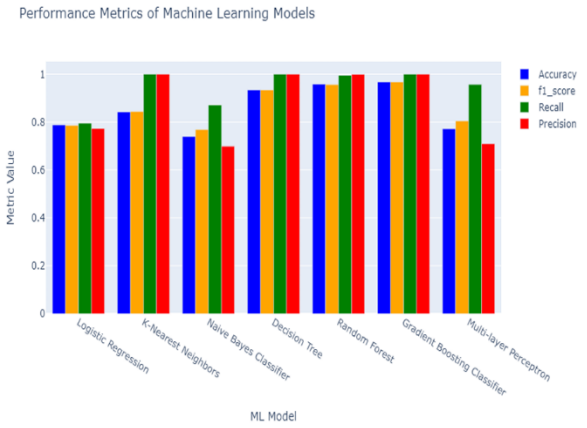

Fig. 5. Perfomance Metrics of Machine Learning Models-Dataset-2

A. Gradient Booster Classifier

The Gradient Boosting Classifier (GBC) imported from scikit-learn outperforms other models in phishing URL de tection due to its combination of efficiency, robustness, and f lexibility.A flowchart is given in fig7 illustrating the sequen tial process starting with the initialization of the model and progressing through sequential training of weak learners. It includes steps for optimizing the model through gradient de scent, combining predictions, and evaluating the loss

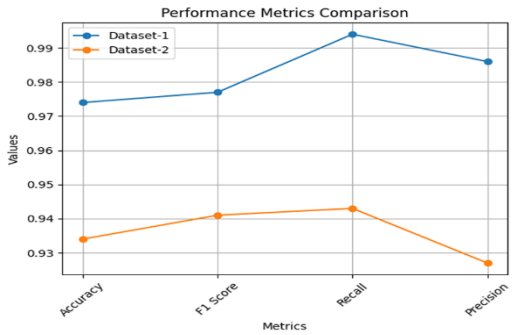

Fig. 6. Perfomance Metrics of Gradient Booster on chosen Datasets

function. Additionally, it incorporates aspects such as hyperparameter tuning, regularization, and scalability considerations to ensure optimal performance. Scikit-learn's implementation of GBC is highly optimized, offering a balance between memory usage and computational speed, which is crucial for processing large datasets efficiently. Additionally, GBC's built-in regu larization techniques help mitigate overfitting, ensuring that the model generalizes well to unseen data. Scikit-learn pro vides user-friendly tools for hyperparameter tuning, allowing practitioners to fine-tune the model's parameters effectively. Furthermore, the library offers methods to analyze feature importance, aiding in the selection of relevant features for phishing URL detection. The compatibility of scikit-learn with other Python libraries and its extensive documentation support further contribute to its effectiveness in phishing URL detection.

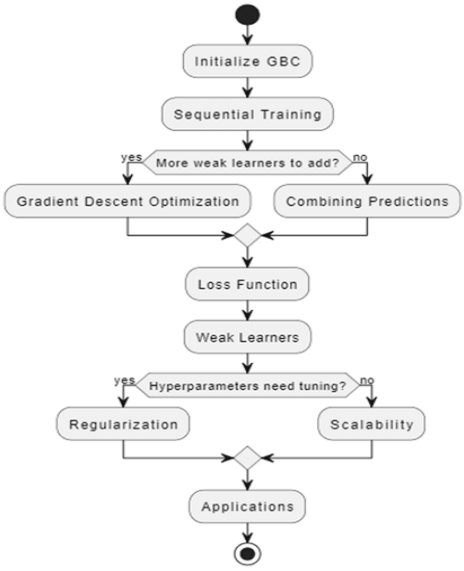

Fig. 7. Gradient Booster Classifier

5. **Conclusion and future works**

To sum up, models with high recall, accuracy, precision, and F1-score have shown promising outcomes when machine learning techniques are applied for phishing URL identifi cation. We have demonstrated how machine learning can effectively identify between phishing and legitimate URLs by using varied datasets and classification techniques. The efficiency of these models signifies their capacity to augment cybersecurity protocols by effectively detecting suspicious URLs and alleviating any hazards. However, there are still issues and restrictions to be resolved even if machine learning-based techniques have demonstrated significant effectiveness in phishing URL detection.Since phishing attempts are dynamic and fraudsters' strategies are always evolving,the detection models must be continuously improved and adjusted. Furthermore, the supervised learning approach's dependence on labelled datasets could make it difficult to precisely capture the subtleties of phishing URLs and generalize it to the unobserved data. Future studies in machine learning-based phishing URL de tection can include developing learning frameworks for real time threat adaptation, integrating interpretability techniques to improve model transparency, exploring unsupervised learning approaches for anomaly detection without labelled data, and further refining feature engineering techniques to extract more informative attributes from URLs. They also involve work ing with security ecosystems to create an all-encompassing defense strategy and enhancing model robustness against ad versarial attacks. By tackling these issues, we can improve cybersecurity risks mitigation in the face of dynamic threats and build the efficacy and robustness of phishing URL detec tion systems.

Refrences

[1] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari, and S. R. K. Joga, "Phish- ing detection system through hybrid machine learning based on url," IEEE Access, vol. 11, pp. 36 805–36 822, 2023.

[2] P. L. Indrasiri, M. N. Halgamuge, and A. Mohammad, "Robust ensemble ma- chine learning model for filtering phishing urls: Expandable random gradient stacked voting classifier (erg-svc)," IEEE Access, vol. 9, pp. 150 142–150 161, 2021.

[3] M. Sameen, K. Han, and S. O. Hwang, "Phishhaven—an efficient real time ai phishing urls detection system," IEEE Access, vol. 8, pp. 83 425–83 443, 2020.

[4] Y. A. Alsariera, V. E. Adeyemo, A. O. Balogun, and A. K. Alazzawi, "Ai meta- learners and extra-trees algorithm for the detection of phishing websites," IEEE access, vol. 8, pp. 142 532–142 542, 2020.
[5] S. Al-Ahmadi, A. Alotaibi, and O. Alsaleh, "Pdgan: Phishing detection with generative adversarial networks," IEEE Access, vol. 10, pp. 42 459–42 468, 2022.
[6] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. U. Haq, K. Saleem, and M. H. Faheem, "A deep learning-based phishing detection system using cnn, lstm, and lstm-cnn," Electronics, vol. 12, no. 1, p. 232, 2023.
[7] E. A. Aldakheel, M. Zakariah, G. A. Gashgari, F. A. Almarshad, and A. I. Alzahrani, "A deep learning-based innovative technique for phishing detection in modern security with uniform resource locators," Sensors, vol. 23, no. 9, p. 4403, 2023. vi
[8] R. Yang, K. Zheng, B. Wu, C. Wu, and X. Wang, "Phishing website detec- tion based on deep convolutional neural network and random forest ensemble learning," Sensors, vol. 21, no. 24, p. 8281, 2021.
[9] M. A. Adebowale, K. T. Lwin, and M. A. Hossain, "Intelligent phishing detec- tion scheme using deep learning algorithms," Journal of Enterprise Information Management, vol. 36, no. 3, pp. 747–766, 2023.
[10] Madasamy, N. S., Eldho, K. J., Senthilnathan, T., & Deny, J. (2023). A Novel Back-Propagation Neural Network for Intelligent Cyber-Physical Systems for Wireless Communications. *IETE Journal of Research*, *70*(2), 1361–1373. https://doi.org/10.1080/03772063.2023.2173669.
[11] E. Chand, "Phishing website detector," https://www.kaggle.com/datasets/ eswarchandt/phishing-website-detector/data, 2020, kaggle Dataset
[12] Dr.Eldho K J, A Framework for Developing Measurement Systems Based on Software Metrics Using SMMT(January10,2019),SRN: https://ssrn.com/abstract=3317410 or http://dx.doi.org/10.2139/ssrn.331740
[13] Hannousse, Abdelhakim; Yahiouche, Salima (2021), "Web page phish ing detection", Mendeley Data, V3, doi: 10.17632/c2gw7fy2j4.3