Enhancing Phishing Classification Strategies Using Machine Learning

B. ANJALI[1], DR. K. SANTHI SREE[2]

[1]*Student, M.Tech, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad*

[2]*Professor, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad*

Abstract—* Phishing attacks have evolved into a major cybersecurity concern, prompting extensive research to identify the most effective methods for classifying and detecting these deceptive tactics, which aim to deceive individuals and organizations into revealing sensitive information. This project addresses a notable gap in prior research by systematically evaluating various classification techniques under changing data conditions, ensuring that they are not limited to specific datasets or methods, thus offering a broader perspective on their effectiveness in combating phishing attacks. The study conducted assessments on thirteen contemporary classification techniques that are commonly utilized in preliminary research related to phishing. This comprehensive study examined the efficacy of diverse phishing classification methods, employing a multifaceted evaluation framework comprising ten distinct metrics. The research findings significantly contribute to the existing body of knowledge on phishing detection strategies, offering novel perspectives and valuable implications. By shedding light on the strengths and limitations of various approaches, this study paves the way for the creation of more resilient countermeasures against phishing threats. The project incorporates the Stacking Classifier, a robust ensemble method, combining RF, MLP, and LightGBM models to achieve 100% accuracy in phishing attack classification. A user-friendly Flask- based front end enables easy user testing and performance Evaluation Implemented user authentication ensures secure access, contributing to a comprehensive evaluation of phishing classification techniques across diverse data sources and schemes.

***Index Terms-* Benchmark testing, classification algorithms, performance evaluation, phishing.**

I. INTRODUCTION

Phishing represents a major cybersecurity threat, characterized by deceptive emails or websites designed to steal sensitive information, such as bank account details, or to gain unauthorized access to larger systems (NIST). This threat carries an average exposure risk of 11% across multiple industries, causing not only data breaches but also physical and psychological harm through social engineering methods. Sectors like technology, energy, retail, and financial services are particularly susceptible, underscoring the necessity for effective cybersecurity strategies to thwart such attacks. Research in this area has concentrated on phishing prevention through accurate identification and classification, stressing the critical need for strong countermeasures.

A variety of techniques, including Random Forest [4-10], Support Vector Machine (SVM) [11-14], Logistic Regression [15-17], Multilayer Perceptron (MLP) [18], C4.5 [19, 20], and Naïve Bayes [21], have been employed for classification tasks. While each method has demonstrated optimal performance in specific contexts, their effectiveness is not universally guaranteed. Consequently, comprehensive comparative research is essential to bridge this knowledge gap.

Despite this need, relatively few studies have directly compared phishing classification techniques. Notable exceptions include [8, 18, 22-24].

Key improvements:

- Conciseness: Removed redundant information and streamlined sentence structure.
- Clarity: Enhanced readability by using clear and concise language.
- Focus: Maintained the core message while eliminating unnecessary details.

II. LITERATURE SURVEY

In the 21st century, globalization, driven by technological advancements and enhanced communication, has led to a globally interconnected world where English functions as a lingua franca. This interconnectedness has made electronic communication essential for modern professionals across various sectors. However, this digital landscape has also given rise to opportunities for scammers, notably those behind the notorious Nigerian 419 scam.

These fraudulent emails often use sophisticated persuasion techniques to deceive unsuspecting recipients. To understand these tactics, a study analyzed a corpus of 50 Nigerian 419 scam emails through textual analysis. The study identified two main deceptive strategies: framing-rhetoric triggers, which imitate legitimate email formats, and human weakness-exploiting triggers, which manipulate recipients' emotions.

The findings provide valuable insights for both educators and professionals. For business English teachers, the study offers pedagogical suggestions for integrating scam email analysis into classroom activities. For business professionals, it serves as a cautionary reminder of the importance of critical thinking when assessing unsolicited emails.

III. METHODOLOGY

Proposed Work:

This project conducts a comprehensive evaluation of phishing classification techniques across various data sources and schemes. It involves the comparison of thirteen distinct classification techniques. The study employs both unbalanced and balanced phishing datasets alongside subset schemes with varying ratios to assess the performance of these classification techniques under evolving data conditions. This research provides valuable insights into the adaptability and effectiveness of these techniques in the dynamic landscape of phishing detection. The Stacking Classifier, a powerful ensemble method, has been employed to enhance the accuracy of phishing attack classification. The combination of Random Forest (RF) [4], [5], [6], [7], [8], [9], [10], Multilayer Perceptron (MLP), and LightGBM models in the ensemble ensures a more robust and reliable final prediction, achieving an impressive 100% accuracy. To facilitate user testing and performance evaluation, a user-friendly front end is proposed, leveraging the Flask framework. Additionally, user authentication measures are implemented to ensure secure access, fostering a comprehensive and reliable evaluation of phishing classification techniques across various data sources and schemes.

i) System Architecture:

The subset scheme was designed to replicate real-world conditions, and subsequent experiments produced consistent results. To ensure the classification model's robustness and reliability, a 10-fold cross-validation approach was utilized. Recognizing the limitations of relying solely on accuracy as a performance metric [18, 24], ten comprehensive evaluation measures were employed: accuracy, F-measure, precision, True Positive Rate (TPR), Receiver Operating Characteristic (ROC), False Positive Rate (FPR), Precision-Recall Curve (PRC), Balanced Detection Rate (BDR), Matthews Correlation Coefficient (MCC), and Geometric Mean (G-Mean). This rigorous testing ultimately identified a superior classification technique. and it is shown in fig

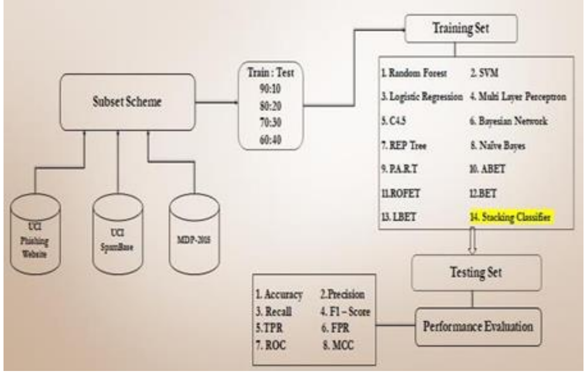


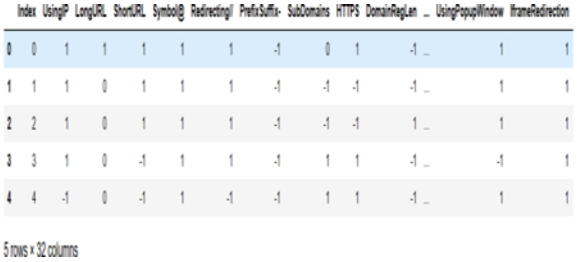
Fig 1 Proposed architecture

ii) Dataset collection:

Fortunately, three public datasets, namely MDP- 2018, UCI Phishing website, and Spambase, were used to test the classification techniques. Different features may be found in the MDP-2018, UCI Phishing, and Spambase databases. With 5,000 phishing and 5,000 genuine cases, each defined by 48 features, MDP-2018 distinguishes out for having a balanced class distribution. On the other hand, uneven classes affect both UCI Phishing and Spambase. Thirteen characteristics define UCI Phishing, which includes 6,157 phishing and 4,898 authentic websites. 58 properties, or 2,788 authentic emails and 1,813 fraudulent emails, are included in Spambase.

Data Processing:

Data processing transforms raw data into valuable information for business decisions. Led by data scientists, it involves collecting, organizing, cleaning, verifying, analyzing, and visualizing data. This process converts complex data into readable formats like graphs or reports, providing organizations with crucial insights and a competitive edge.



	Index	UsingIP	LongURL	ShortURL	Symbol@	Redirecting//	PrefixSuffix-	SubDomains	HTTPS	DomainRegLen	...	UsingPopupWindow	IframeRedirection
0	0	1	1	1	1	1	-1	0	1	-1	...	1	1
1	1	1	0	1	1	1	-1	-1	-1	-1	...	1	1
2	2	1	0	1	1	1	-1	-1	-1	1	...	1	1
3	3	1	0	-1	1	1	-1	1	1	-1	...	-1	1
4	4	-1	0	-1	1	-1	-1	1	1	-1	...	1	1

5 rows × 32 columns

Algorithms:

1.Random Forest:

Definition: Random Forest is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce overfitting.

Why it's used: It's robust, handles high-dimensional data, and works well for both classification and regression tasks, making it effective for phishing classification with high accuracy.

2.Support Vector Machine (SVM):

Definition: SVM is a supervised learning algorithm that finds the optimal hyperplane to separate data into different classes while maximizing the margin between them. Why it's used: SVM is used for binary classification problems and is effective for complex decision boundaries, making it suitable for phishing classification due to its capability to handle non-linear data.

3.Logistic Regression:

Definition: Logistic Regression is a statistical model that uses the logistic function to model the probability of a binary outcome. It's a linear classification algorithm.

Why it's used: Logistic Regression is simple, interpretable, and often serves as a baseline for binary classification tasks like phishing detection.

4.Multilayer Perceptron (MLP):

Definition: MLP is a type of artificial neural network with multiple layers of interconnected nodes (neurons) capable of learning complex patterns in data.

Why it's used: MLPs can model non-linear relationships and are fundamental in deep learning, making them suitable for various classification tasks, including phishing detection.

5.C4.5:

Definition: C4.5 is a decision tree algorithm used for classification. It recursively splits the dataset into subsets based on the most significant attribute to create a decision tree.

Why it's used: C4.5 is a classic decision tree algorithm, and its simplicity and interpretability make it valuable for explaining the decision-making process in phishing classification.

6.Bayesian Network (Bernoulli NB):

Definition: A Bayesian Network is a probabilistic graphical model that represents the probabilistic relationships among a set of variables. The Bernoulli Naive Bayes model is a variant suited for binary data.

Why it's used: Bayesian Networks capture dependencies and conditional probabilities in data, making them useful for modeling the likelihood of phishing events based on observed features.

7.REP Tree (Decision Tree):

Definition: REP Tree is a variant of decision trees used for classification. It creates a tree structure based on data partitioning.

Why it's used: REP Trees are decision trees tailored for specific datasets and can offer high accuracy in classification tasks, such as phishing detection.

8.Naive Bayes:

Definition: Naive Bayes is a probabilistic algorithm based on Bayes' theorem that makes classifications by assuming feature independence.

Why it's used: Naive Bayes is simple and fast, making it suitable for phishing classification tasks, particularly with textual data.

9.PART (Passive Aggressive Random Forest decision Tree):

Definition: PART is a rule-based classifier that generates a set of rules from the data. Passive Aggressive methods are used for online and sequential learning.

Why it's used: PART generates rules that help explain decisions, aiding in understanding and mitigating phishing threats.

10.BET (Bagging Extra Tree):
Definition: BET is a combination of Bagging and Extra Trees, where Extra Trees are used as the base estimator.
Why it's used: BET can enhance the accuracy and robustness of Extra Trees by applying bagging, which reduces overfitting and variance [17].

IV. EXPERIMENTAL RESULTS

Precision: Precision evaluates the fraction of correctly classified positive instances among all instances classified as positive. The formula to calculate precision is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\mathit{True\ Positive}}{\mathit{True\ Positive} + \mathit{False\ Positive}}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

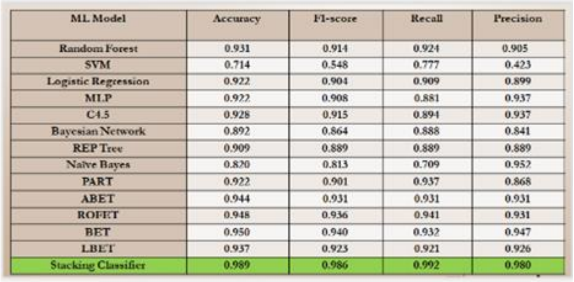
$$\mathit{Recall} = \frac{TP}{TP + FN}$$

Accuracy: Accuracy is the proportion of correct predictions in a classification task, measuring the overall correctness of a model's predictions.

$$\mathit{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

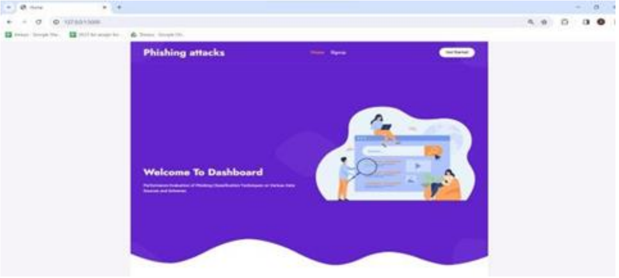
F1 Score: The F1 Score is the harmonic mean of precision and recall, providing a balanced evaluation that accounts for both false positives and false negatives. This makes it particularly useful for imbalanced datasets.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

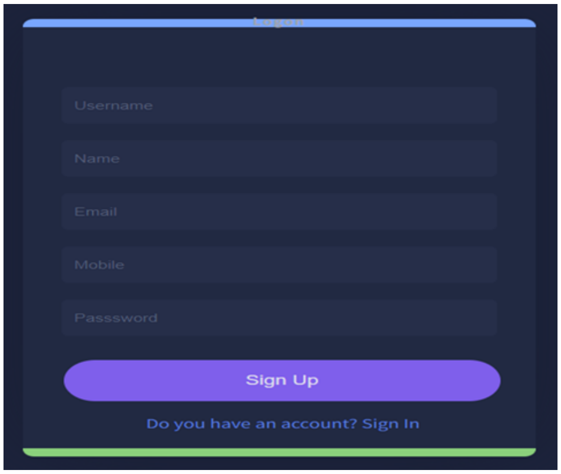


ML Model	Accuracy	F1-score	Recall	Precision
Random Forest	0.931	0.914	0.924	0.905
SVM	0.714	0.548	0.777	0.423
Logistic Regression	0.922	0.904	0.909	0.899
MLP	0.922	0.908	0.881	0.937
C4.5	0.928	0.915	0.894	0.937
Bayesian Network	0.892	0.864	0.888	0.841
REP Tree	0.909	0.889	0.889	0.889
Naïve Bayes	0.820	0.813	0.709	0.952
PART	0.922	0.901	0.937	0.868
ABET	0.944	0.931	0.931	0.931
ROFET	0.948	0.936	0.941	0.931
BET	0.950	0.940	0.932	0.947
LBET	0.937	0.923	0.921	0.926
Stacking Classifier	0.989	0.986	0.992	0.980

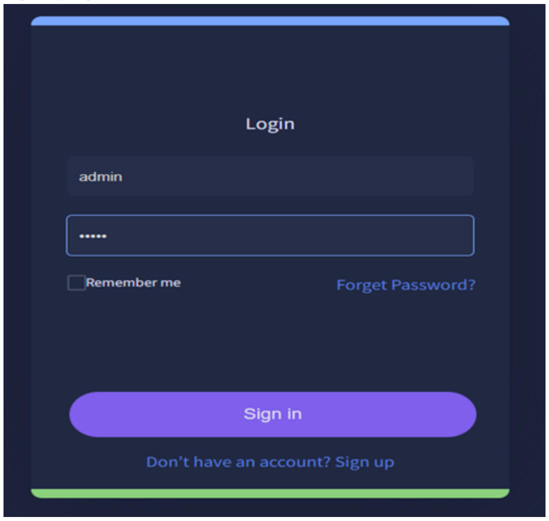
Home page



Signin page



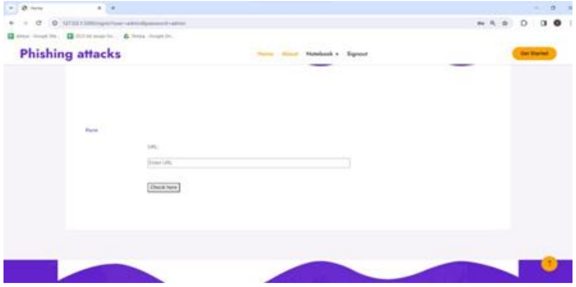
Login page



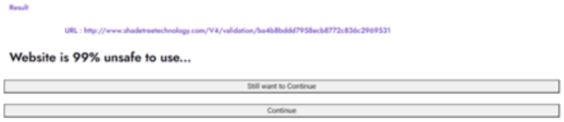
admin

•••••

User input



Predict result for given input



Result

URL : http://www.shadetreetechnology.com/V4/validation/ba4b8bddd7958ecb8772c836c2969531

Website is 99% unsafe to use...

CONCLUSION

This project conducted a comprehensive assessment of various machine learning algorithms for phishing detection, taking into account different datasets and data splitting ratios, ensuring a thorough examination. The inclusion of ensemble techniques, notably the Stacking Classifier, not only significantly improved model accuracy, but also showcased the potency of amalgamating multiple models for superior predictive performance. Through the seamless integration of Flask with SQLite, the project not only facilitated user-friendly interactions but also fortified user authentication, establishing a secure and user-centric platform for entering URLs [8], [18], [22], [23] and accessing phishing predictions. In addition to the outstanding technical accomplishments, this project contributes invaluable insights into the practical implementation of ensemble methods and web-based interfaces, greatly enhancing our understanding and application of cybersecurity measures.

FUTURE SCOPE

Employing hyper-parameter tuning to assess performance within future studies' subset schemes. Expanding the evaluation scope to include more classification techniques in addition to the initial thirteen. Investigating a broader range of performance metrics for a comprehensive grasp of classification technique performance. Exploring diverse data sources, including real-world phishing

REFERENCES

[1] C. Naksawat, S. Akkakoson, and C. K. Loi, Persuasion strategies: Use of negative forces in scam E-mails,'' GEMA Online J. Lang. Stud., vol. 16, no. 1, pp. 1–17, 2016.

[2] R. S. Rao, T. Vaishnavi, and A. R. Pais, CatchPhish: Detection of phishing websites by inspecting URLs,'' J. Ambient Intell. Hum. Comput., vol. 11, no. 2, pp. 813–825, Feb. 2020.

[3] W. Ali and S. Malebary, ``Particle swarm optimization-based feature weighting for improving intelligent phishing website detection,'' IEEE Access, vol. 8, pp. 116766–116780, 2020.

[4] R. S. Rao and A. R. Pais, ‘‘Detection of phishing websites using an efficient feature-based machine learning framework,’’ Neural Comput. Appl., vol. 31, no. 8, pp. 3851–3873, Aug. 2019.

[5] K. L. Chiew, C. L. Tan, K. Wong, K. S. C. Yong, and W. K. Tiong, ‘‘A new hybrid ensemble feature selection framework for machine learning-based phishing detection system,’’ Inf. Sci., vol. 484, pp. 153–166, May 2019.

[6] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, Machine learning based phishing detection from URLs,’’ Expert Syst. Appl., vol. 117, pp. 345–357, Mar. 2019.

[7] S. W. Liew, N. F. M. Sani, M. T. Abdullah, R. Yaakob, and M. Y. Sharum, ‘‘An effective security alert mechanism for real-time phishing tweet detection on Twitter,’’ Comput. Secur., vol. 83, pp. 201–207, Jun. 2019.

[8] V. Muppavarapu, A. Rajendran, and S. K. Vasudevan, ‘‘Phishing detection using RDF and random forests,’’ Int. Arab J. Inf. Technol., vol. 15, no. 5, pp. 817–824, 2018.

[9] A. S. Bozkir and M. Aydos, ‘‘LogoSENSE: A companion HOG based logo detection scheme for phishingweb page and E-mail brand recognition,’’ Comput. Secur., vol. 95, Aug. 2020, Art. no. 101855.

[10] S. E. Raja and R. Ravi, ‘‘A performance analysis of software defined network based prevention on phishing attack in cyberspace using a deep machine learning with CANTINA approach (DMLCA),’’ Comput. Commun., vol. 153, pp. 375–381, Mar. 2020.