ANALYSIS OF PHISHING CLASSIFICATION METHODS & VERIFYING THE AUTHENTICITY OF URLS

V. Sowmya Devi[a*], Ch. Niranjan Kumar[a], P. Punitha[c]
[a] Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India
[c] Department of CSE(DS), Vignana Bharathi Institute of Technology, Hyderabad, Telangana, India

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 paper 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. It subjected them to ten diverse performance measures, aiming to provide a comprehensive understanding of their capabilities. This paper 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. Secure access is ensured by implemented user authentication, which helps to support an exhaustive assessment of phishing classification strategies across various schemes and data sources.

Keywords: Benchmark testing, classification algorithms, performance evaluation, phishing.

******Author for Correspondence*** E-mail: sowmya.v@sreenidhi.edu.in

INTRODUCTION

Phishing is a dangerous threat to online security that has been identified by the National Institute of Standards and Technology as an attempt to obtain sensitive information such as account numbers or gain access to key electronic systems through fraudulent requests made via email or websites. The typical gamble of being helpless against this assault across assorted areas is 11% [1]. Phishing is a socially designed assault that truly hurts people and associations [2]. The corporate sectors are Innovation, Energy or Utilities, Retail and Financial Management. These clubs are particularly vulnerable to phishing. To forestall these assaults, network safety measures should be carried out [3]. A few examinations on phishing evasion have been directed, with one zeroing in on its ID and grouping.

Ordering approaches include random forests [4], [5], [6], [7], [8], [9], [10], support vector machines (SVM) [11], [12], [13], [14], Logistic Regression [15], [16], [17], Multilayer Perceptron (MLP) [18], C4.5 [19] and [20], Naive Bayes [21]. Each works best for its application. The arrangement method's outcomes don't need to be all around relevant. Therefore, similar review should be led to close this hole.

A couple of exploration have looked at phishing order methods, including [8], [18], [22], [23], and [24]. This correlation research is isolated into four significant segments: phishing, dataset type, execution assessment, and systems. [8], [18], [22], [23], and [24] accepted their information from a phishing site and URL, though [24] utilized crude messages from Apache Spam Assassin and Nazario. The most widely recognized exhibition assessments are exactness, accuracy, and F-measure. The most normally used methods are Random Forest, SVM, and Naïve Bayes. This relative exploration has a hole, which is what existing methodologies mean for particular public datasets, both adjusted and uneven.

Shockingly, this review depends on the presentation assessment of the order method while utilizing a particular lopsided dataset for various phishing sorts. This is practically identical to the methodologies utilized in research that didn't analyze the classification procedures. Vaitkevicius and Marcinkevicius [18] investigated two adjusted and one imbalanced dataset. It was accounted for that they beat before examinations. Gana and Abdulhamid [23] just utilized uneven public datasets and exhibited that arrangement execution changes as indicated by the subset strategy. This review is based on various tests that failed to demonstrate how execution evaluation affects the strategies used to adjust different subsets of dataset plans. Some simply point out that the exhibition's impact on commonly used schedules such as 90:10, 80:20, 70:30, and 60:40 is slight. Additionally, execution evaluation and ordering methods are mandated by accompanying measurements: accuracy, F-measure, precision, true positive rate (TPR), receiver operating characteristic (ROC), false positive rate (FPR), precision recall curve. (PRC), Matthews correlation coefficient (MCC), balanced detection rate (BDR), and geometric mean. A subset of each pattern is shown to influence the presentation score of the grouping method on both matched and heterogeneous datasets. This will in general fundamentally improve and debase the presentation of particular subgroups.

LITERATURE SURVEY

Globalization in the twenty-first century fundamentally affects the world because of significantly further developed innovation and correspondence, permitting everybody required to have equivalent admittance to an overall market and data exchange by means of English. Therefore, electronic correspondence has turned into a norm for the present worldwide experts in all areas, who work consistently before computerized screens. Now and again, these experts might get Nigerian 419 trick messages in which con artists request that casualties make settlements ahead of time for monetary prizes that won't ever appear. These messages contain very much created circumstances where influence methodologies are intermixed. Subsequently, the casualty who is helpless against the deal is bound to answer and eventually lose cash. Thus, the ongoing review led a text based investigation on a corpus of 50 Nigerian 419 trick messages to research language components as far as influence strategies utilized by fraudsters as a convincing power to satisfy their open objectives of draws and double dealings. The review [2] distinguished two critical kinds of misleading systems that are utilized in mix: outlining way of talking triggers, masked as the ordinary type of electronic correspondences, and human shortcoming taking advantage of triggers, expected to mix beneficiaries' feelings. At long last, the paper incorporates not just instructive ideas for business English instructors while carrying out homeroom exercises, yet additionally alerts for both pre-endlessly experienced business experts on the most proficient method to decipher obscure email messages with intense wariness.

There are various enemy of phishing methods that influence source code-based elements and outsider administrations to distinguish phishing locales. These procedures have inadequacies, one of which is their failure to oversee drive-by downloads. They additionally utilize outsider administrations to identify phishing URLs, which creates setbacks for the characterization cycle. Thus, in this study [4], we present CatchPhish, a lightweight program that predicts URL authenticity without requiring the client to visit the site. The proposed procedure use the Irregular backwoods classifier to arrange hostnames, whole URLs [4, 13, 21, 26], Term Frequency-Inverse Document Frequency (TF-IDF) qualities, and phishing-indicated phrases from dubious URLs. The proposed model, which utilized exclusively TF-IDF qualities from our dataset, achieved a precision of 93.25%. The trial with TF-IDF and hand-made highlights accomplished a significant precision of 94.26% on our dataset and 98.25%, 97.49% on benchmark datasets, which is essentially higher than the current pattern models.

Web phishing attacks have become progressively refined as of late, provoking clients to lose trust in web based business and online organizations. Different strategies and methods in view of a boycott of phishing sites are utilized to identify phishing locales [8, 9, 10, 11, 13]. Sadly, the quick extension of innovation has brought about the rise of additional modern procedures for building sites that draw in buyers. Hence, current boycott based frameworks neglect to recognize the latest and recently sent off phishing sites, for example, zero-day phishing sites. A few late investigations have utilized ML calculations to recognize phishing sites and use them as an early advance notice framework to identify such dangers. Nonetheless, in most of these methodologies, the fundamental site attributes have been picked in light of human experience or recurrence examination. This work [5] proposes insightful detection of phishing sites using molecular group rationalization-based highlight weighting that acts on the IDs of phishing sites. The proposed approach involves leveraging particle swarm optimization (PSO) to productively weight different site boundaries to detect phishing sites with a higher degree of accuracy. Specifically, we use the proposed PSO-based site highlight weighting to distinguish the importance of specific site components in detecting phishing and genuine sites. Exploratory results show that the proposed PSO-based saliency weighting further improves the accuracy of ML model characterization, true positive and negative proportions, and false positive and false negative proportions, while at the same time reducing the phishing used We showed that localization reduces website salience.

Phishing is a type of digital attack that tricks unsuspecting Internet-based customers into revealing sensitive information such as usernames, private keys, government-backed pension numbers, and visa numbers. Assailants delude Web clients by acting like a reliable or credible site page to get individual data. There have been different enemy of phishing arrangements introduced to far, including boycotts and whitelists, as well as heuristic and visual comparability based calculations, yet web buyers keep on being tricked into unveiling basic data on phishing sites. In this review [6], we introduce a new ordering model using heuristic highlights collected from URLs, source code, and external controls to address the deficiencies of current adversaries in phishing techniques. Our model is evaluated using eight different ML strategies, including random forest (RF) calculations [4], [5], [6], [7], [8], [9], [10] exceeded. 99.31% accuracy. Investigations continued using several (symmetrical and lateral) irregular Timberland classifiers to determine the best classifier to distinguish between phishing sites. Principal Component Analysis Random Forest (PCA-RF) outperforms all gradient RFs with 99.55% accuracy. Additionally, to examine how effective external managers are in characterizing unsafe locations, we evaluated models with and without outsider-based factors. We also compared our findings with sample models (CANTINA and CANTINA+). Our proposed method outperforms these methods while identifying zero-day phishing attacks.

This study presents an alternative factor decision structure for ML-based phishing detection framework known as Hybrid Ensemble Feature Selection (HEFS) [7]. The first step in HEFS is to create a subset of essential components using a proprietary cumulative distribution function gradient (CDF-g) method, which is then transformed into data processing groups to create optional element subsets. An optional subset of elements is used in his second step to create a set of pattern highlights using the skill stimulus collection. In general, the test results show that HEFS performs best when combined with a random forest classifier. This meter can accurately detect 94.6% of phishing and legitimate websites, but requires only 20.8% initial emphasis. In another study, random forest pattern highlighting (total 10 48). PART classifier was used. HEFS is also comparable to another notable phishing dataset from the University of California, Irvine (UCI) vault. Therefore, HEFS becomes a very attractive and reasonable element selection technique for ML-based phishing location frameworks.

METHODOLOGY

i) Proposed Work:

This research evaluates phishing classification approaches using various data sources and schemes. It includes a comparison of thirteen different classification techniques. The review utilizes both imbalanced and adjusted phishing datasets, as well as subset plans with various proportions, to assess the exhibition of these arrangement approaches under changing information circumstances. This study reveals insight into the adaptability and efficiency of different systems in the steadily changing phishing recognition climate. The Stacking Classifier, a sophisticated ensemble algorithm, was used to improve 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. Furthermore, user authentication methods are established to provide secure access, allowing for a thorough and trustworthy evaluation of phishing categorization techniques across several data sources and schemes.

ii) System Architecture:

The subset plot was contrived to reflect the genuine circumstances, and the analysis yielded comparable outcomes when applied thusly. To confirm that the characterization model created is magnificent and trustworthy, a 10-overlay cross-approval technique was utilized. It isn't prudent to depend simply on precision to assess execution [18], [24]. This brought about the reception of eleven execution assessment measurements, including accuracy, F-measure, precision, TPR, ROC, FPR, PRC, BDR, MCC, and G-mean. At last, a characterization procedure that performed well in these tests was distinguished, as displayed in Figure 1.

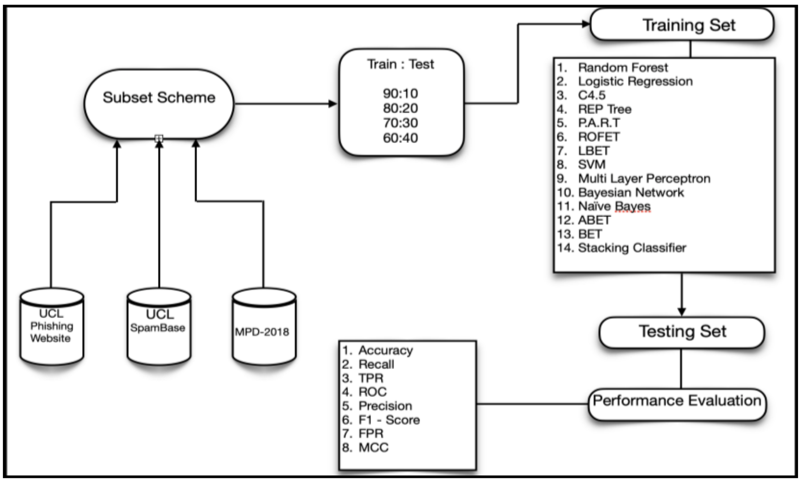
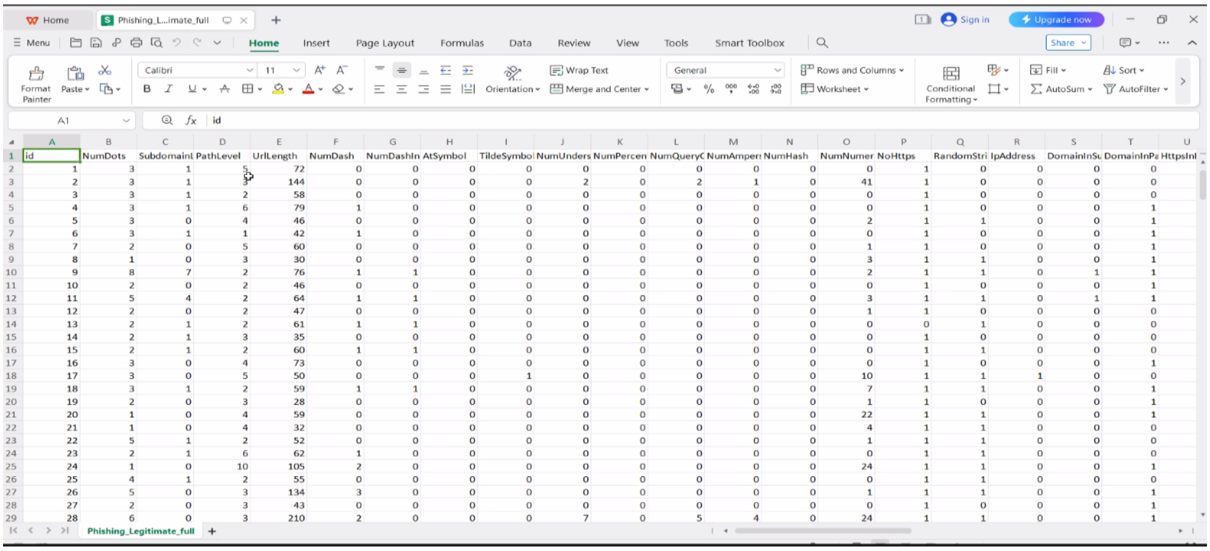


Fig. 1: *Proposed architecture*

iii) Data set collection :

Fortunately, three accessible datasets were used to evaluate the ordering method: MDP-2018, UCI phishing site, and Spam base. The MDP-2018 dataset has a moderate class distribution, but the UCI phishing sites and spam-based datasets have an imbalanced distribution. [33] There are a total of 5,000 phishing and trusted websites. MDP-2018 has 48 highlights, while UCI Spam base has 58 items with good records, including 2,788 genuine messages and 1,813 fake messages. The UCI phishing site consists of 31 components containing records from 6,157 phishing sites and 4,898 real sites.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	id	NumDots	Subdomainl	PathLevel	UrlLength	NumDash	NumDashIn	AtSymbol	TildeSymbo	NumUnders	NumPercen	NumQueryC	NumAmpers	NumHash	NumNumer	NoHttps	RandomStri	IpAddress	DomainInSu	DomainInPa	HttpsInl
2	1	3	1	5	72	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
3	2	3	1	3	144	0	0	0	0	2	0	2	1	0	41	1	0	0	0	0	
4	3	3	1	2	58	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
5	4	3	1	6	79	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	
6	5	3	0	4	46	0	0	0	0	0	0	0	0	0	2	1	1	0	0	1	
7	6	3	1	1	42	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	
8	7	2	0	5	60	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	
9	8	1	0	3	30	0	0	0	0	0	0	0	0	0	3	1	1	0	0	1	
10	9	8	7	2	76	1	1	0	0	0	0	0	0	0	2	1	1	0	1	1	
11	10	2	0	2	46	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	
12	11	5	4	2	64	1	1	0	0	0	0	0	0	0	3	1	1	0	1	1	
13	12	2	0	2	47	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	
14	13	2	1	2	61	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	
15	14	2	1	3	35	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
16	15	2	1	2	60	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	
17	16	3	0	4	73	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	
18	17	3	0	5	50	0	0	0	1	0	0	0	0	0	10	1	1	1	0	0	
19	18	3	1	2	59	1	1	0	0	0	0	0	0	0	7	1	1	0	0	1	
20	19	2	0	3	28	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	
21	20	1	0	4	59	0	0	0	0	0	0	0	0	0	22	1	1	0	0	1	
22	21	1	0	4	32	0	0	0	0	0	0	0	0	0	4	1	1	0	0	0	
23	22	5	1	2	52	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	
24	23	2	1	6	62	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	
25	24	1	0	10	105	2	0	0	0	0	0	0	0	0	24	1	1	0	0	1	
26	25	4	1	2	55	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	
27	26	5	0	3	134	3	0	0	0	0	0	0	0	0	1	1	1	0	0	1	
28	27	2	0	3	43	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	
29	28	6	0	3	210	2	0	0	0	7	0	5	4	0	24	1	1	0	0	1	

Fig. 2: **UCI phishing dataset**

iv) Data Processing:

Data processing entails converting raw data into useful information for businesses. Data scientists typically process data by gathering, organizing, cleaning, verifying, analyzing, and translating information into understandable formats such as graphs or papers. Data processing can be done in three ways: manually, mechanically, and electronically. The goal is to increase the value of information and facilitate decision-making. This allows organizations to improve their operations and make more timely strategic decisions.

Automated data processing technologies, such as computer programming, play an important role in this. It can help transform enormous amounts of data, including big data, into useful insights for quality management and decision-making.

v) Feature selection:

Feature selection is the method involved with distinguishing the most steady, non-excess, and pertinent elements for use in model creation. As data sets fill in amount and assortment, it is basic to deliberately decrease their size. The basic role of component determination is to expand the exhibition of a prescient model while diminishing the computational expense of demonstrating.

Feature selection, one of the important parts of component design, is a demonstration of selecting the main highlights to be considered in ML calculations. Include selection techniques are used to limit the amount of information elements by eliminating repetitive or redundant elements and focusing on highlights that are generally valuable to the ML model. There are significant benefits to making saliency decisions much earlier, rather than relying on ML models to determine which elements are important.

vi) Algorithms:

The following algorithms are considered for phishing classification using various data sources.

Name of the Algorithm	Definition	Importance
Random Forest	It is an ensemble learning strategy that uses different selection trees for estimation. Further improve accuracy and reduce overfitting by building a series of decision trees and prioritizing their predictions.	It is robust, can handle high-dimensional data, and is useful for both classification and regression problems. In the case of phishing classification, it can provide a high level of accuracy [4], [5], [6], [7], [8], [9], [10].
s.Support Vector Machine (SVM)	It is a managed learning strategy that decides the best hyperplane to split information into classes while expanding the edge between them.	It is used for binary classification issues, and it works especially well with complex decision boundaries. It is commonly used in phishing classification because of its capacity to handle nonlinear data [11], [12], [13], and [14].
Logistic regression	It is a fact-based model that uses logistic functions to calculate the probabilities of double outcomes. This is purely a grouping calculation.	It is a basic, interpretable approach that is frequently used as a baseline for binary classification tasks such as phishing detection [15], [16], and [17].
Multilayer Perceptron (MLP)	It is a kind of artificial neural network made out of various layers of interconnected nodes (neurons) fit for learning confounded designs in information	They are utilized because of their capacity to represent non-linear relationships, and they are a key component of deep learning. They may perform a variety of categorization tasks, including phishing detection [18].
C4.5	It is a decision tree strategy utilized in classification. To produce a choice tree, it recursively parcels the dataset into subsets in light of the main property.	It is a traditional decision tree algorithm whose simplicity and interpretability make it useful for describing the decision-making process in phishing categorization [19, 20].
Bayesian Network (Bernoulli NB)	It is a probabilistic graphical model that portrays the probabilistic relationships between's factors. The Bernoulli Naive Bayes model is reasonable for double information.	They can capture dependencies and conditional probabilities in data, making them ideal for predicting the likelihood of future occurrences based on observed features.

Decision Tree	It is a type of decision tree used for classification. It builds a tree structure based on data partitions.	Decision trees designed for specific datasets and can provide high accuracy in classification applications such as phishing detection.
Naive Bayes	It is a probabilistic calculation that applies Bayes' hypothesis. It performs orders by assuming that elements are free, a "naive" yet as often as possible compelling presumptions.	It is a simple and quick text classification method that is well-suited for phishing classification tasks, particularly when dealing with textual data [21].
PART (Passive Aggressive Random Forest decision Tree)	It is a rule-based classifier that generates a set of rules based on the data. Passive Aggressive methods are typically used for online and sequential learning.	It can generate rules that explain why a particular decision was made, which can be useful for understanding and mitigating phishing threats.
ABET (AdaBoost ExtraTree)	It is an ensemble learning algorithm that combines Extra Trees with AdaBoost. Extra Trees are a variation of Random Forest.	AdaBoost with Extra Trees can improve classification performance by combining the strengths of both algorithms. It can be particularly effective for handling imbalanced datasets [29].
ROFET (Random Forest ExtraTree)	It combines Random Forest with Extra Trees, which are random decision trees	It combines the robustness of Random Forest with the variance reduction of Extra Trees, potentially improving overall classification accuracy.
BET (Bagging ExtraTree)	It is a combination of Bagging and Extra Trees, where Extra Trees are used as the base estimator	It can enhance the accuracy and robustness of Extra Trees by applying bagging, which reduces overfitting and variance [17].
LBET (Logistic Gradient ExtraTree)	It is a hybrid model combining logistic regression and Extra Trees.	It can provide a balance between the interpretability of logistic regression and the power of Extra Trees, making it useful for explaining and classifying phishing instances.
Stacking Classifier (RF + MLP with LightGBM)	It is an ensemble technique that combines multiple base models (Random Forest and MLP) using a meta-model (LightGBM).	It leverages the strengths of different algorithms, potentially improving overall classification accuracy and robustness for phishing detection.

EXPERIMENTAL RESULTS

Precision: Precision estimates the extent of precisely characterized cases or tests among those classified as certain.

$$Precision = \frac{TP}{TP + FP} \qquad (1)$$

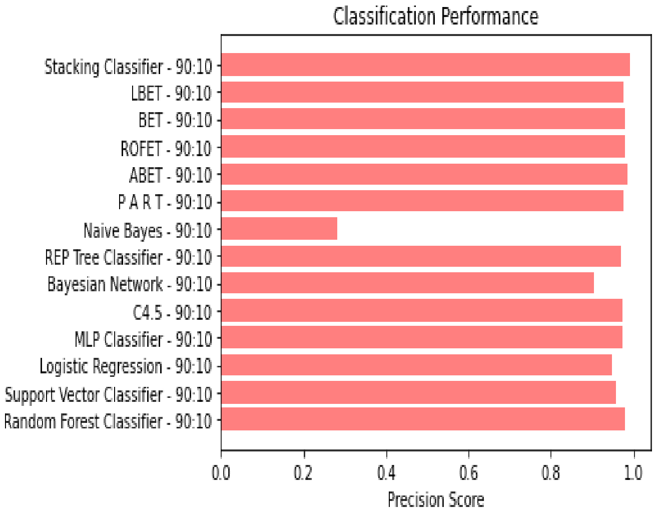


***Fig. 3:** Precision comparison graph*

Recall: Recall is an ML metric that evaluates a model's capacity to perceive all occasions of a given class. It is the proportion of accurately anticipated positive perceptions to add up to real up-sides, which gives data on a model's fulfillment in gathering instances of a particular class.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

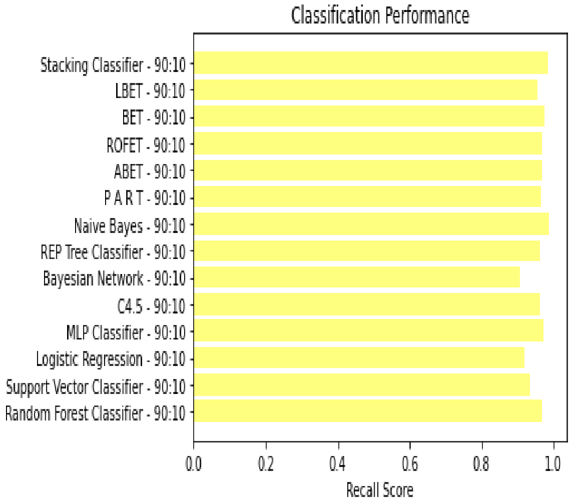


***Fig. 4:** Recall comparison graph*

Accuracy: Accuracy is characterized as the extent of right forecasts in a grouping position, which estimates a model's general accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

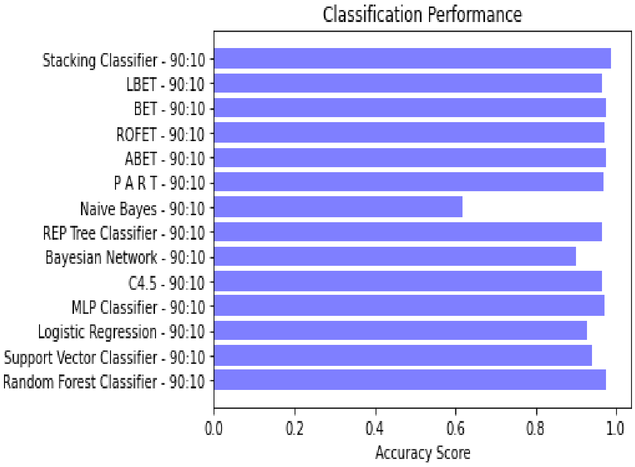


Fig. 5: *Accuracy graph*

F1 Score: The F1 Score, the symphonious mean of accuracy and recall, is suitable for imbalanced datasets since it accounts for false positives and negatives.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (4)$$

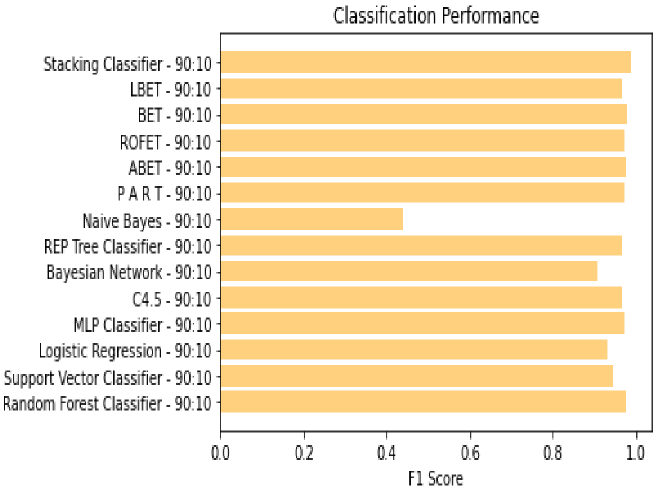


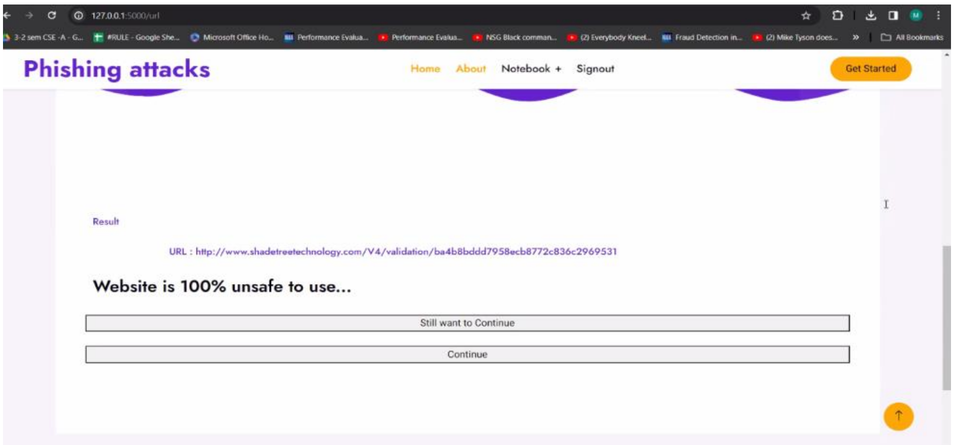
Fig. 6: *F1Score*

Table 1: *Performance Evaluation*

ML Model	Accuracy	Fl-score	Recall	Precision
Random Forest	0.931	0.914	0.924	0.905
SVM	0.714	0.548	0.777	0.423
Logistic	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	0.892	0.864	0.888	0.841
REP Tree	0.909	0.889	0.889	0.889
Naive B ayes	0.82	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	0.95	0.94	0.932	0.947
LBET	0.937	0.923	0.921	0.926
Stacking	0.989	0.986	0.992	0.98

Application Development

From the various Machine learning models, stacking algorithm as performed well. So, an application is developed by using stacking model to check whether the given website is prone to phishing attack or whether it is safe. Here first we need to enter the URL, based on parameters it displays whether the website is safe or unsafe.

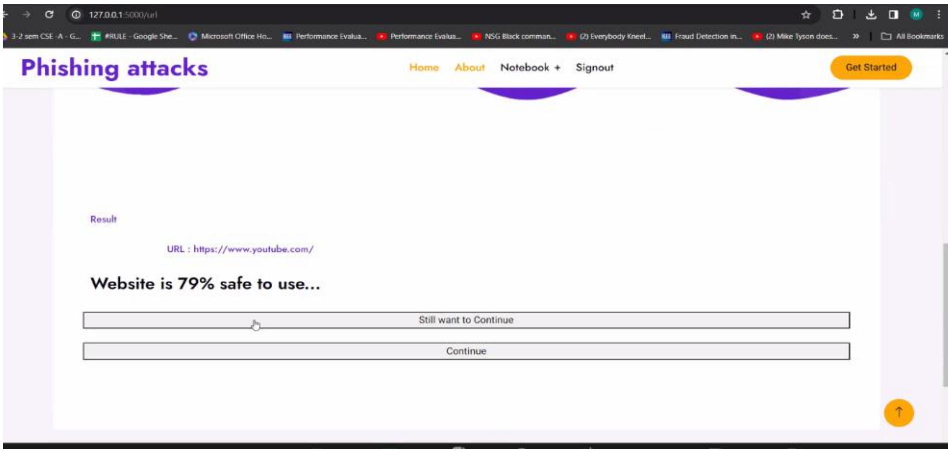


Result

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

Website is 100% unsafe to use...

Fig. 7: *Unsafe URL*



Result

URL : https://www.youtube.com/

Website is 79% safe to use...

Fig. 8: *Unsafe URL*

CONCLUSIONS

This research 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 and accessing phishing predictions. In addition to the outstanding technical accomplishments, this paper 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 datasets and industry-specific data, to assess classification technique performance in varied contexts.

REFERENCES

1. Proofpoint. (Jun. 2022). 2021 State of the Phish. [Online]. Available: https://www.proofpoint.com/sites/default/files/gtd-pfpt-us-tr-state-of-thephish-2020.pdf
2. R. Abdillah, Z. Shukur, M. Mohd, T. S. M. Z. Murah, I. Oh and K. Yim, "Performance Evaluation of Phishing Classification Techniques on Various Data Sources and Schemes," in *IEEE Access*, vol. 11, pp. 38721-38738, 2023, doi: 10.1109/ACCESS.2022.3225971.
3. Goenka, Richa, Meenu Chawla, and Namita Tiwari. "A comprehensive survey of phishing: mediums, intended targets, attack and defence techniques and a novel taxonomy." *International Journal of Information Security* 23.2 (2024): 819-848.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. A. S. Bozkir and M. Aydos, ‘‘LogoSENSE: A companion HOG based logo detection scheme for phishing web page and E-mail brand recognition,’’ Comput. Secur., vol. 95, Aug. 2020, Art. no. 101855.
12. 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.
13. M. Sameen, K. Han, and S. O. Hwang, ‘‘PhishHaven—An efficient real-time AI phishing URLs detection system,’’ IEEE Access, vol. 8, pp. 83425–83443, 2020.
14. R. S. Rao, T. Vaishnavi, and A. R. Pais, ‘‘PhishDump: A multi-model ensemble based technique for the detection of phishing sites in mobile devices,’’ Pervasive Mobile Comput., vol. 60, Nov. 2019, Art. no. 101084.
15. Y. Ding, N. Luktarhan, K. Li, and W. Slamu, ‘‘A keyword-based combination approach for detecting phishing webpages,’’ Comput. Secur., vol. 84, pp. 256–275, Jul. 2019.
16. A. K. Jain and B. B. Gupta, ‘‘A machine learning based approach for phishing detection using hyperlinks information,’’ J. Ambient Intell. Hum. Comput., vol. 10, no. 5, pp. 2015–2028, May 2019.
17. A. E. Aassal, S. Baki, A. Das, and R. M. Verma, ‘‘An in-depth benchmarking and evaluation of phishing detection research for security needs,’’ IEEE Access, vol. 8, pp. 22170–22192, 2020.
18. P. Vaitkevicius and V. Marcinkevicius, ‘‘Comparison of classification algorithms for detection of phishing websites,’’ Informatica, vol. 31, no. 1, pp. 143–160, Mar. 2020.
19. Y.-H. Chen and J.-L. Chen, ‘‘AI@ntiPhish—Machine learning mechanisms for cyber-phishing attack,’’ IEICE Trans. Inf. Syst., vol. E102.D, no. 5, pp. 878–887, May 2019.
20. Varshney, Gaurav, et al. "Anti-phishing: A comprehensive perspective." *Expert Systems with Applications* 238 (2024): 122199.
21. S. Mishra and D. Soni, ‘‘Smishing detector: A security model to detect smishing through SMS content analysis and URL behavior analysis,’’ Future Gener. Comput. Syst., vol. 108, pp. 803–815, Jul. 2020.
22. M. Karabatak and T. Mustafa, ‘‘Performance comparison of classifiers on reduced phishing website dataset,’’ in Proc. 6th Int. Symp. Digit. Forensic Secur. (ISDFS), Mar. 2018, pp. 1–5.

23. N. N. Gana and S. M. Abdulhamid, ''Machine learning classification algorithms for phishing detection: A comparative appraisal and analysis,'' in Proc. 2nd Int. Conf. IEEE Nigeria Comput. Chapter (NigeriaComputConf), Oct. 2019, pp. 1–8.
24. T. Gangavarapu, C. D. Jaidhar, and B. Chanduka, ''Applicability of machine learning in spam and phishing email filtering: Review and approaches,'' Artif. Intell. Rev., vol. 53, no. 7, pp. 5019–5081, Oct. 2020.
25. S. Priya, S. Selvakumar, and R. L. Velusamy, ''Evidential theoretic deep radial and probabilistic neural ensemble approach for detecting phishing attacks,'' J. Ambient Intell. Hum. Comput., vol. 14, no. 3, pp. 1951–1975, Jul. 2021.
26. P. L. Indrasiri, M. N. Halgamuge, and A. Mohammad, ''Robust ensemble machine learning model for filtering phishing URLs: Expandable random gradient stacked voting classifier (ERG-SVC),'' IEEE Access, vol. 9, pp. 150142–150161, 2021.
27. A. Ozcan, C. Catal, E. Donmez, and B. Senturk, ''A hybrid DNN-LSTM model for detecting phishing URLs,'' Neural Comput. Appl., vol. 35, no. 7, pp. 4957–4973, Aug. 2021.
28. S.-J. Bu and H.-J. Kim, ''Optimized URL feature selection based on genetic-algorithm-embedded deep learning for phishing website detection,'' Electronics, vol. 11, no. 7, p. 1090, Mar. 2022.
29. V. Zeng, S. Baki, A. E. Aassal, R. Verma, L. F. T. De Moraes, and A. Das, ''Diverse datasets and a customizable benchmarking framework for phishing,'' in Proc. 6th Int. Workshop Secur. Privacy Anal., Mar. 2020, pp. 35–41.
30. A. Ihsan and E. Rainarli, ''Optimization of k-nearest neighbour to categorize Indonesian's news articles,'' Asia–Pacific J. Inf. Technol. Multimedia, vol. 10, no. 1, pp. 43–51, Jun. 2021.
31. 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. 142532–142542, 2020.
32. E. Sukawai and N. Omar, ''Corpus development for Malay sentiment analysis using semi supervised approach,'' Asia–Pacific J. Inf. Technol. Multimedia, vol. 9, no. 1, pp. 94–109, Jun. 2020. [33] C. L. Tan, ''Phishing dataset for machine learning: Feature evaluation,'' Mendeley Data, V1, 2018, doi: 10.17632/h3cgnj8hft.1.
33. X.-Y. Lu, M.-S. Chen, J.-L. Wu, P.-C. Chang, and M.-H. Chen, ''A novel ensemble decision tree based on under-sampling and clonal selection for web spam detection,'' Pattern Anal. Appl., vol. 21, no. 3, pp. 741–754, Aug. 2018.
34. M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, ''A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches,'' IEEE Trans. Syst., Man C, Appl. Rev., vol. 42, no. 4, pp. 463–484, Jul. 2012.
35. E. S. Gualberto, R. T. De Sousa, T. P. D. B. Vieira, J. P. C. L. Da Costa, and C. G. Duque, ''From feature engineering and topics models to enhanced prediction rates in phishing detection,'' IEEE Access, vol. 8, pp. 76368–76385, 2020.
36. Alnemari, Shouq, and Majid Alshammari. "Detecting phishing domains using machine learning." *Applied Sciences* 13.8 (2023): 4649.
37. I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Amsterdam, The Netherlands: Elsevier, 2017.
38. T. Saito and M. Rehmsmeier, ''The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,'' PLoS ONE, vol. 10, no. 3, pp. 1–21, Mar. 2015.
39. T. Fawcett, ''An introduction to ROC analysis,'' Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Dec. 2006.
40. A. E. Aassal, L. Moraes, S. Baki, A. Das, and R. Verma, ''Anti-phishing pilot at ACM IWSPA 2018: Evaluating performance with new metrics for unbalanced datasets,'' in Proc. Anti-Phishing Shared Task Pilot 4th ACM IWSPA, 2018, pp. 2–10.
41. H. A. Alshalabi, S. Tiun, and N. Omar, ''A comparative study of the ensemble and base classifiers performance in Malay text categorization,'' Asia– Pacific J. Inf. Technol. Multimedia, vol. 6, no. 2, pp. 53–64, Dec. 2017.
42. N. Japkowicz and M. Shah, Evaluating Learning Algorithms. Cambridge, U.K.: Cambridge Univ. Press, 2011.
43. R. Gowtham and I. Krishnamurthi, ''PhishTackle—A web services architecture for anti-phishing,'' Cluster Comput., vol. 17, no. 3, pp. 1051–1068, Sep. 2014.