AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.5, NO. 4(2)(2025)

COMPREHENSIVE EVALUATION OF PHISHING DETECTION TECHNIQUES ACROSS DIVERSE DATASETS AND CLASSIFICATION MODELS

[1]Supriya Gande, [2] Mrs P.Sunitha(Assistant professor),
[3]Ch. Aravind Kumar(Assistant professor), [4]Dr.N.Satyavathi(Associate professor)
[1,2]CSE, Vaagdevi College of Engineering (Bollikunta, Warangal)

ABSTRACT

Phishing attacks have become a big problem for cyber security, which has led to many studies to find the best ways to find out and categorize these tricks that try to get people and businesses to provide private information. This project fulfills a large gap in previous research by systematically testing various classification methods on data that changes and ensures that they are not limited to certain data sets or methods. This gives a wider view of how well they work to stop phishing attacks. The study dealt with thirteen modern methods of categorization, which are often used in early phishing research. They performed 10 different performance tests to get a complete picture of what they could do. The results of this study contribute to our understanding of phishing classification techniques by transcending what happened in other studies on the same topic. This will help us come up with better ways to protect yourself from phishing threats. Research uses stacking classifier, strong file approach that combines "RF, MLP and LightGBM models to classify 100% phishing attacks correctly". The front end built on the flask that is easy to use is simple for testing and evaluating performance. User verification ensures that users only have access to the system, which helps to explore different phishing categories in a number of data sources and schemes.

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

1. INTRODUCTION

Phishing is a serious danger to cyber security. The National Institute for Standards and Technology says that phishing is when someone sends fake requirements via E -e -mail or website to get sensitive information such as bank account data or access to larger computer systems. The average risk of attack in different industries is 11%. [1]. Phishing is another kind of socially modified attack that hurts people and businesses, whether physically or mentally [2]. The business sector includes technologies, energy or public services, retail and financial services. Phishing is a great threat to these groups. To stop these attacks, we need to use measures in the area of cyber security [3]. There were several studies on how to stop phishing, one of which was focused on searching and sorting phishing e -mails.

There are several ways to classify things such as "random forest [4], [5], [6], [7], [8], [9], [10], support vector machine (SVM) [11], [12], [13], [14], Logistic regression [15], [16], [17], Multilayer perceptron (MLP) [18], C4.5 [19] and [20], and Naïve Bayes [21]". Each of them works well in the situation in which it was used. The results of the categorization method may not be the same in each scenario. A comparative study should be performed to fill in this vacuum.

However, few researches were concerned with how different methods of phishing categorization work, including [8], [18], [22], [23] and [24]. "There are four primary aspects of this comparative research: phishing, type of data file, performance assessment and method used". Data sources for [8], [18], [22], [23] and [24] came from the phishing site and URL. [24] It uses raw e -maily from Apache spamassassin and Nazario. The most important evaluation of performance is the "accuracy, precision, and F-measure". The most common methods are "naive Bayes, SVM and random forest". In this

comparative study, there is a void in terms of how current methods affect different public data sets, both balanced and unbalanced.

Interestingly, this study examines how well the classification approach works on a given unbalanced data file for different types of phishing. It is like what research that these methods did not compare. "Vaitkevicius and Marcinevicius" [18] used two balanced "data files and one that was not". It was said that they got better results than in an earlier comparison. "Gana and Abdulhamide" [23] used only unbalanced public data files and it has been shown that the classification performance varies depending on how the subset is located. This study is based on a number of other studies that could not show how the performance evaluation affects the methods used to sort different groups of data sets schemes. Some were just talking about how little this performance affected conventional schemes, such as 90:10, 80:20, 70:30 and 60:40. Evaluation and classification methods are also limited by measurements such as "accuracy, F-measuring, accuracy, real positive speed (TPR), operational characteristics of the receiver (ROC), false positive rate (FPR), accuracy-receptation curve (PRC), Matthews coeficient (MCC), balanced detection rate (BDR). It has been shown that each subset of the scheme in balanced and unbalanced data files affects how well the classification approach works. This often causes the presentation of different subgroups to go up and down a lot.

2. LITERATURE REVIEW

Globalization in the 21st century is having a massive influence on the world as a result of the fact that technology and communication have become so much improved to an extent that a global market is accessible to everyone whereas communication can be done using English language. Owing to this, the electronic communication is currently an aspect of the daily occurrences by the professionals who belong to any profession and operate their digital screens. The luckless ones of this caliber find themselves in a situation whereby these professionals receive Nigerian 419 scam mail, an email message by fraudulent individuals attempting to get individuals to pay advance money that never arrives. These emails are very good illustrations how various persuasive tools can go along. Due to this, the victim that is likely to accept the offer will most likely respond and lose cash. Thus, the present study analyzed 50 Nigerian 419 fraud email messages with the help of textual analysis in order to understand how the language was used by scammers as the means of persuasion in order to convey their messages and convince people to give them money. The study [2] identified two broad categories of tricks that are commonly used in conjunction with one another: “framing-rhetoric triggers that appear to setup no more than a common electronic communication, and human weakness-leveraging triggers, that are intended to cause people to have some kind of feeling”. Finally, the paper provides the business English teachers with ideas on the manner of utilizing classroom activities and at the same time cautions the business people, who are either new to the profession or have been in the line of business, to be very cautious and read carefully the messages in unknown e-mails.

The number of methods of preventing phishing that utilise source code-based services and features of other companies to locate the phishing sites is quite big. Among the pitfalls with these ways, they are not compatible with drive by downloads. They use the third party services to locate phishing URLs as well, and this hinders the process of categorization. Thus, in this paper [4] we propose implementation of one lightweight program name CatchPhish which can determine an existence of a real URL without even accessing it. Random Forest classifier uses hostname, the full URL [4, 13, 21, and 26], “Term Frequency-Inverse Document Frequency (TF-IDF) features and phish-hinted

phrases in the suspect URL to distinguish between the data". The model using a single TF-ID function on our data had an accuracy of 93.25 percent. TF-ID with manually created functions reached an accuracy of 94.26 percent on our data set and 98.25 percent and 97.49 percent in the Benchmark data set. This is much better compared to the accuracy of basic models.

The customers have lost confidence in e-commerce and online services since phishing attacks through the web were evolving continuously "in the past few years. A number of systems and methods exist" that identify phishing websites by using a list of known phishing sites [8, 9, 10, 11, and 13]. Unfortunately, technology is dynamic to the extent that new and more enhanced methods of building websites in order to attract consumers have returned. Due to this, such of these blacklist-based techniques cannot detect the most recently discovered and launched phishing websites and the likes of these zero-day phishing websites. Other more recent research papers have been applying ML at identifying phishing websites and using the websites as an early warning mechanism of detecting such risks. The majority of these means, however, select the most important site attributes with the help of the human experience or the frequency of web site properties. This article [5] provides an idea of using weighing functions based on particle optimization to increase phishing identification. The proposed algorithm is based on the application "Optimization of particle swarms (PSO)" to assign appropriate weighing to different websites in searching for phishing sites. The proposed weighing features of the PSO -based website is used to distinguish the difference between different features on the web in terms of their importance in telling the difference between "phishing and legitimate sites". Experimentally, it was found that the suggested scheme weighing PSO -based features has turned machine learning models into a much better tool for classifying phishing websites in terms of real and negative rates and false positive and negative rates, while weighing PSO -based features used by a smaller number of functions provided by the site to achieve the same.

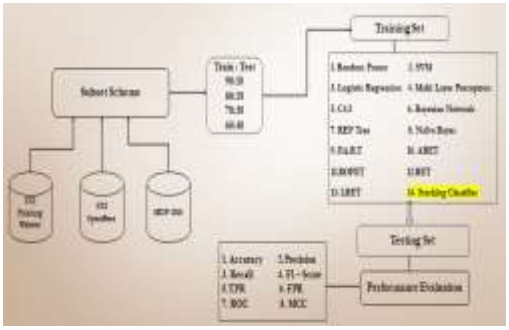
Phishing is an example of a cyberattack that lures oblivious individuals to share personal information like "login, password, social security number, or credit card number". Attackers can fool Internet users into providing them with their personal information by making an Internet site appear to be another one that is trustworthy or seems to be a legitimate site. Many anti-phishing tools have been proposed; these include blacklists and whitelists, heuristic methods and ones based on visual similarity. Fixed sites are still getting people to go to phishing sites to provide personal data. In the current research [6], we will propose a new model of categorization that uses heuristic properties of "URL, source code and third-party services to overcome the issues of existing techniques of anti-phishing". We used eight ML algorithms to test our model. "Random Forest (RF) method [4], [5], [6], [7], [8], [9], [10] performed best here", and had an accuracy in this case of 99.31%. To find the optimal "random forest classifier to use in the detection of phishing websites, they were tested using other random forest class (orthogonal and oblique). The best of all oblique Random Forests (oRFs) which is Random Forest (PCA-RF) achieved the highest accuracy rate which was 99.55 percent". "We also tested our model with or without third-party features and observed how third party services assisted us in finding questionable sites". We also studied the extent to which our results were] =9lec our proposed strategy performed better than these and was also able to detect phishing attacks on that very day.

This study proposes an alternate method to select features "to use in a phishing detection

system based on machine learning". It is known as the "Hybrid Ensemble Feature Selection (like the HEFS)" [7]. In step 1 of HEFS, primary feature sets are obtained with a new method called the "Cumulative Distribution Function gradient (CDF-g)". As a result, the subset of the subset is to insert the subset of the secondary functions into a data failure file. In the second phase, a set of basic elements is obtained on the basis of a set of failure function of functions using subset of secondary elements. General test results show that the use of HeF with random forests classifier is the most prolific. The basic resolution correctly distinguishes 94.6 percent of phishing and authentic websites on only 20.8 percent of data in the Congress. When running another test on Random Forest to test the top 10 baseline characteristics, it performed superior to SVM when using 48 features [11], [12], [13], [14], "NaiveBayes, C4. 5, JRip, and PART classifiers". HEFS also reports good results when the corpus is tested with another widely used phishing corpus found in the UCI repository. Owing to this, the HEFS is a useful and desirable method that is used in selecting features in phishing detection systems that employ the use of machine learning.

3. MATERIALS AND METHODS

This study examines the methods of phishing categorization on various data sources and schemes. It compares thirteen comparison outlooks to categorize things. This study implements both unbalanced and balanced phishing datasets and old ratios setup to ensure the greatest effectiveness of this kind of classification. This study provides us with the valuable evidence of how adaptable and practical those strategies can be even in the realm of rapidly changing phishing detection. To make the categorization of phishing attacks more potent, we employed a powerful file-like algorithm (Stacking Classifier). "RF [4], [5], [6], [7], [8], [9], [10], MLP and LightGBM" make a final prediction by using majority rules as ensemble, and are highly accurate and reliable with 100 percent accuracy. I would recommend a user-friendly front end using the Flask framework so that the users could be tested, and performance could be checked with ease. There are also user authentication steps that are used to ensure that there is access security. This helps to provide a comprehensive and reliable evaluation of phishing categories to schemes across the severity of data.



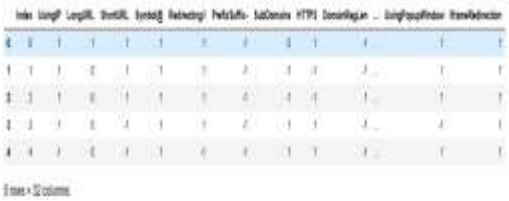
"Fig.1 Proposed Architecture"

The technique of subset was customized to the reality on the ground and the following experiment that was conducted had comparable results. "We employed a 10-fold cross-validation approach to ensure that the model of categorization that we worked out is decent and trustworthy". The use of the accuracy as an indicator of performance is not good at all [18], [24]. It gave way to the application of 10 measures of performance assessments, including the "accuracy, F-measure, precision, and TPR, ROC, FPR, PRC, BDR, MCC, and G-Mean". At last, one classification approach was identified that performed well on these tests all. It is displayed in the form shown in Fig. 1.

a) Dataset Collection:

It is good that three public data sets such as: MDP-2018, UCI Phishing and Spambase website have been used to test categorization techniques. UCI Phishing and Spambase databases have disproportionality in classes, but this is not the case with MDP-2018. There are 5,000 phishing websites and 5,000 actual websites [33]. MDP-2018 consists of 48 functions and UCI Spamse consists of 58

functions and 2788 right and 1813 fake e-mails are present as records. The UCI Phishing website has 31 data functions 6 157 Phishing websites and 4,898 authentic websites.



"Fig 2 UCI phishing dataset"

b) Data-Processing:

In processing data, raw data is being converted to useful information to organizations. Data scientists normally treat data through collection, organization, cleaning, checking, analyzing, and translating them into formats that other people can comprehend e.g., graphs or words. There are three methods through which you will process data; with a hand, a machine, or a computer. This will be to enhance the usefulness of information and to assist people to make decisions. This aids in the improved performance of organizations and also making important decisions in time. A large part of this is computer programming and other automated tools of data processing. It would enable you to interpret much data, in fact immense data, in order to make the best decisions and govern quality.

c) Feature selection:

The selection of functions refers to the process of identifying the most constructive, most useful and unique features to be used to create a model. Since the number and amount of data sets continues to grow, it is necessary to systematically reduce the size of these data sets. The basic reason for choosing elements is to perform the performance of the predictive model and reduce modeling costs.

The important aspect of feature engineering is feature selection. It is making decisions about the most important features to be utilised as part of the ML algorithms. The feature selection approaches reduce the level of input variables by eliminating those characteristics that are not instrumental or very similar to the other characteristics. This reduces the size of the set of features to the most helpful in the ML model. The key reasons to pick features in advance rather than simply letting the ML model discover the features that are most important are

d) Algorithms:

1. "Random Forest":

- Definition: Random Forest is a method of learning that uses a set of decision -making trees for prediction. It grows the forest of decision -making trees and aggregates their predictions to make them the least accurate and less likely.

- Why it's used: Random Forest is powerful, can deal with high dimensions data and is applicable to classification and regression tasks. It can be very sensible to apply in the process of classifying phishing [4], [5], [6], [7], [8], [9], [10].

"Fig 3 Random forest"

2. "Support Vector Machine (SVM)":

- "Definition: SVM is a supervised machine learning algorithm which finds the optimum hyperplane to" segregate two or more datasets and also separates a space between each other and in a manner that the room is as wide as possible.

- "Why it's used: SVM is used when dealing with binary classification problems", where it is hard to interpret the decision lines. It is commonly employed in categorizing phishing since it can process the data which is not straightforward [11], [12], [13], [14].

“Fig 4 SVM”

3. “Logistic Regression”:

- Definition: The logistical regression concerns a statistical model that uses a logistics function to determine the probabilities with which the binary result will have a positive result. It is a linear classification algorithm.

- “Why it's used: Logistic regression is simple and broadly employed” as an entry point to the binary classification problems such as phishing detection [15], [16], and [17].

“Fig 5 Logistic regression”

4. “Multilayer Perceptron (MLP)”:

- Definition: LP belongs to the group of artificial neural networks consisting of multiple layers of interconnected nodes (neurons), which can be trained to recognize complex input relationships.

- Why it's used: MLPs incorporate non-linear connections, which makes them an element of deep learning. It is possible to assign this type of labor to categorize a lot of other jobs, such as identifying phishing emails [18].

“Fig 6 MLP5. C4.5”:

- “Definition: C4.5 is one of the varieties of decision trees” algorithms that help sort things. It repeatedly divides the dataset into smaller groups based on the most crucial characteristic in order to construct a decision tree.

- “Why it's used: C4.5 is a classic decision tree” approach and effortless to comprehend and utilize. This renders it applicable in the description of the phishing categorization process [19, 20].

“Fig 7 C4.5”

6. “Bayesian Network (Bernoulli NB)”:

- “Definition: The Bayesian Networks is a type of graphical model which displays the likelihood of the relationship between various variables”. Bernoulli Naive Bayes model is one of the variations that performs with binary data.

- “Why it's used: The Bayesian Networks” are able to indicate the relationship between points of data and the probability of one occurring given that another one occurs. This will be useful in determining your likelihood of phishing acts as you observe it.

“Fig 8 Bayesian network”

7. “REP Tree (Decision Tree)”:

- “Definition: REP Tree is also a type of decision tree with an application in categorization”. It produces tree like structure with a branching of data.

- “Why it's used: REP Trees are decision trees” that are constructed on some specific sets of data. They can be very precise when it comes to such roles as phishing detection.

“Fig 9 REP tree”

8. “Naive Bayes”:

- “Definition: Naive Bayes is a probabilistic approach and it applies the Bayes theorem”. It generates the classes on the assumption that the characteristics are independent of each other, and that the assumption is a naive one though occasionally useful.

- “Why it's used: Naive Bayes is an easy and fast method of text sorting”, and thus they are fit in doing the phishing classification work, especially when the data is in text form [21].

“Fig 10 Naïve bayes”

9. “PART (Passive Aggressive Random Forest decisionTree)”:

- “Definition: PART is a decision making rule based classifier”. Most people employ passive-aggressive way of studying online and in batches.

- Why it's used: PART may stipulate that as regards why a particular choice is made. It may assist people to comprehend and also deal with phishing assaults.

“Fig 11 PART”

10. “ABET (AdaBoost ExtraTree)”:

- Definition: ABET is a type of ensemble learning that employs two methods Extra Trees and AdaBoost. Extra Trees is a variation of Random Forest.

- Why it's used: AdaBoost with Extra Trees might be used to work better at categorization since the process can incorporate the most effective aspects of both approaches to make the most out of both. It may be rather helpful when one works with an unbalanced data set [29].

"Fig 12 ABET"

11. "ROFET (Random Forest ExtraTree)":

- Definition: ROFET is used with random forest and extra trees, random decision trees.

- Why it's used: ROFET brings together the power of the Random Forest and the capacity of the Extra Trees to minimize the variance, which could lead to an overall improvement in the classification accuracy.

"Fig 13 ROFET"

12. "BET (Bagging ExtraTree)":

- Definition: BET consists of two methods, which are Bagging and Extra Trees. As a basis estimator, Extra Trees are used.

- Why it's used: BET can potentially make Extra Trees more accurate and stable by using bagging, which decreases overfitting and variance [17].

"Fig 14 BET"

13. "LBET (Logistic Gradient ExtraTree)":

- Definition: LBET is Ensemble of logistic regression and Extra Trees.

- Why it's used: LBET can assist with explaining, identifying the cases of phishing as it combines the interpretability aspect of logistic regression, with the potency of Extra Trees.

"Fig 15 LBET"

14. "Stacking Classifier (RF + MLP with LightGBM)":

- Definition: There is a method of combining a number of base models (like Random forest and MLP) into a single model using a meta-model (LightGBM) with stacking.

- Why it's used: The idea of stacking is based on the application of the most efficient elements of a range of algorithms in the process of phishing identification to determine accuracy and reliability in general.

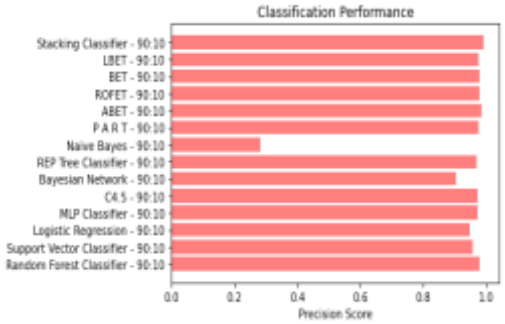
"Fig 16 Stacking classifier"

4. EXPERIMENTAL RESULTS

"Precision": The accuracy determines the number of positively identified samples or instances corrected from all those that have been marked as positive. The equation to be used to find accuracy is therefore:

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

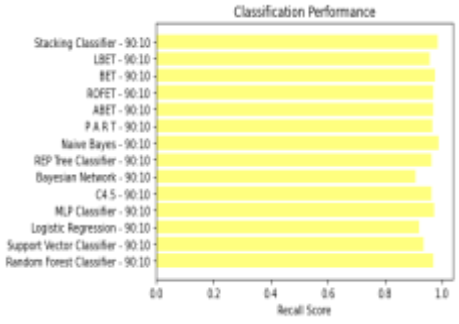
$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$



"Fig 17 Precision comparison graph"

"Recall": Recall has several more or less technical definitions, but in machine learning it is the model's ability to find all the relevant examples of some classes. It is given by the division of properly predicted positives into overall real positives. This will help you an intuitive understanding of how effectively the model can find an instance of one level of this type.

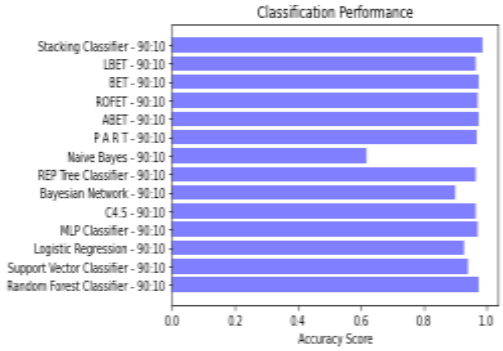
$$Recall = \frac{TP}{TP + FN}$$



"Fig 18 Recall comparison graph"

"Accuracy": Accuracy is a fraction of the exact predictions of the classification task; It indicates to what extent the predictions provided by the model are correct.

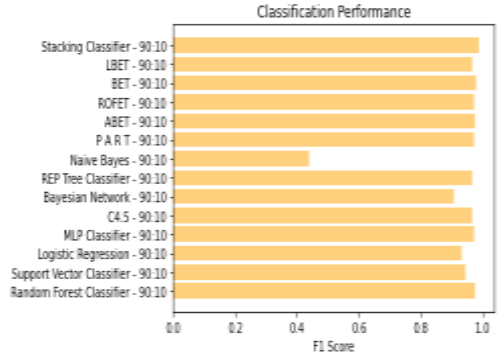
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$



"Fig 19 Accuracy graph"

"F1 Score": The F1 score is a harmonious average of accuracy and appeal. It is a degree of moderation that considers both false positive and negative values and is therefore suitable for non -homogeneous data.

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

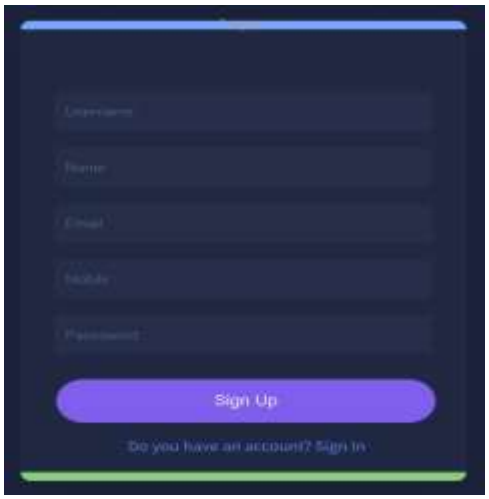


"Fig 20 F1Score"

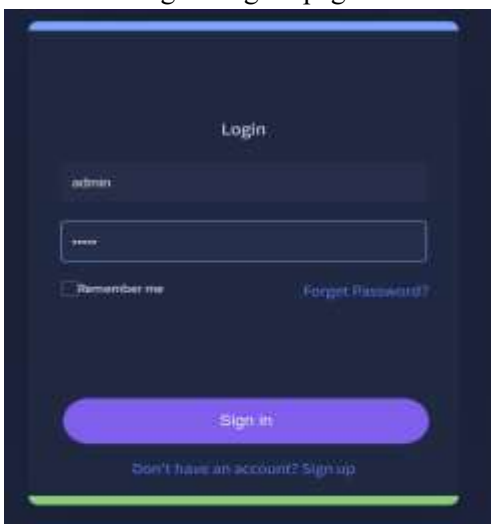
ML Model	Accuracy	Fscore	Recall	Precision
Random Forest	0.931	0.914	0.924	0.905
SVM	0.714	0.648	0.777	0.623
Logistic Regression	0.922	0.904	0.909	0.899
MLP	0.922	0.908	0.881	0.937
C4.5	0.918	0.915	0.894	0.937
Bayesian Network	0.892	0.864	0.888	0.841
REP Tree	0.909	0.889	0.889	0.889
Naive Bayes	0.838	0.813	0.769	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	0.958	0.948	0.952	0.947
LBET	0.937	0.923	0.921	0.926
Stacking Classifier	0.968	0.956	0.963	0.949

"Fig 21 Performance Evaluation"

"Fig 22 Home page"



"Fig 23 Signin page"



"Fig 24 Login page"

"Fig 25 User input"

"Fig 26 Predict result for given input"

5. CONCLUSION

The study involved an extensive analysis of various machine-learning models of phishing detection through a variety of datasets and data splitting proportions to ensure the research has been as comprehensive as it could have been. Completing the model with ensemble techniques and in particular a Stacking Classifier, not not only lifted the model to very high accuracy levels, but also demonstrated the advantages of using a large number of models in order to obtain improvements in model predictions. The project facilitated the involvement of the users in using the program with ease and enhanced the user-authentication, which allowed the friendly integration of Flask to SQLite. This made it safe and easy to enter URLs [8], [18], [22], [23] and obtain phishing predictions. This project is not just a project with huge technological accomplishments, it also provides us with a lot of information concerning the application of ensemble techniques and the use of web based interfaces in the real world. This improves our knowledge and application of cybersecurity actions in a better way.

Measuring performance on future investigations using hyper-parameters with the exception of tweaking on subset schemes. There is also incorporating new methods and means of categorizing in the assessment in addition to first thirteen. Considering a broader range of performance indicators to have a complete idea about the effectiveness of classification methods. To understand how well the methods of categorization apply in various scenarios the use of various types of data, including real-world phishing datasets or data within specific industries can be used [18, 23].

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] 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.

[3] M. A. Pitchan, S. Z. Omar, and A. H. A. Ghazali, ‘‘Amalan keselamatan siber pengguna internet terhadap buli siber, pornografi, e-mel phishing dan pembelian dalam talian (cyber security practice among internet users towards cyberbullying, pornography, phishing email and online shopping),’’ Jurnal Komunikasi, Malaysian J. Commun., vol. 35, no. 3, pp. 212–227, Sep. 2019.

[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] C. L. Tan, K. L. Chiew, K. S. C. Yong, S. N. Sze, J. Abdullah, and Y. Sebastian, ‘‘A graph-theoretic approach for the detection of

phishing webpages,'' Comput. Secur., vol. 95, Aug. 2020, Art. no. 101793.
[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.
[34] 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.
[35] 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.
[36] 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.
[37] H. Zhang, G. Liu, T. W. S. Chow, and W. Liu, ''Textual and visual contentbased anti-phishing: A Bayesian approach,'' IEEE Trans.

Neural Netw., vol. 22, no. 10, pp. 1532–1546, Oct. 2011.
[38] 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.
[39] 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.
[40] T. Fawcett, ‘‘An introduction to ROC analysis,’’ Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Dec. 2006.
[41] 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.
[42] 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.
[43] N. Japkowicz and M. Shah, Evaluating Learning Algorithms. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[44] R. Gowtham and I. Krishnamurthi, ‘‘PhishTackle—A web services architecture for anti-phishing,’’ Cluster Comput., vol. 17, no. 3, pp. 1051–1068, Sep. 2014.