International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

Phishing Detection using Base Classifier and Ensemble Technique

Mithilesh Kumar Pandey[1], Rekha Pal[2], Saurabh Pal[3], Arvind Kumar Shukla[4], Manish Ranjan Pandey[5], Shantanu Shahi[6]
[1,2,3]Department of Computer Applications
VBS Purvanchal University
Jaunpur, India
e-mail: mithileshkumarmca@gmail.com, peehupal08@gmail.com, drsaurabhpal@yahoo.co.in
[4,5]School of Computer Science & Application
IFTM University
Moradabad, India
e-mail: arvindshukla.india@gmail.com, manishmrp123@gmail.com
[6]Department of Computer Science & Engineering
Ambalika Institute of Management & Technology
Lucknow, India
e-mail: shahi261193@gmail.com

Abstract—Phishing attacks continue to pose a significant threat in today's digital landscape, with both individuals and organizations falling victim to these attacks on a regular basis. One of the primary methods used to carry out phishing attacks is through the use of phishing websites, which are designed to look like legitimate sites in order to trick users into giving away their personal information, including sensitive data such as credit card details and passwords. This research paper proposes a model that utilizes several benchmark classifiers, including LR, Bagging, RF, K-NN, DT, SVM, and Adaboost, to accurately identify and classify phishing websites based on accuracy, precision, recall, f1-score, and confusion matrix. Additionally, a meta-learner and stacking model were combined to identify phishing websites in existing systems. The proposed ensemble learning approach using stack-based meta-learners proved to be highly effective in identifying both legitimate and phishing websites, achieving an accuracy rate of up to 97.19%, with precision, recall, and f1 scores of 97%, 98%, and 98%, respectively. Thus, it is recommended that ensemble learning, particularly with stacking and its meta-learner variations, be implemented to detect and prevent phishing attacks and other digital cyber threats.

Keywords- Phishing, Machine learning, Ensemble, Meta-learning, Bagging, Confusion matrix.

I. INTRODUCTION

The growing acceptance and recognition of data innovations have led to an increase in the number of electronic arrangements available through the Internet. Research indicates that electronic monetary exchanges, online gaming, and entertainment are among the most popular arrangements with a large number of users. The goal is to make online services more convenient and widely available for daily use [1]. However, the widespread availability and accessibility of these digital services on the internet also expose them to cyber threats since there are no standard control measures for the internet. Cyber-attacks present serious risks and vulnerabilities for both the digital services and their users, with data and financial losses being among the most severe consequences [2].

A common example of cyber-attacks is the website phishing attack, where cybercriminals create fraudulent websites to deceive unsuspecting users and steal their sensitive information for illicit purposes. This type of cybercrime is a major concern for internet security, with severe impacts on both web users and electronic businesses. Website phishing attacks are a pervasive form of fraud that involves the creation of fake websites designed to look like legitimate ones, with the intention of tricking unsuspecting users into divulging their personal information [3].

In the fourth quarter of 2021, the Anti-Phishing Working Group reported a record-high of 316,747 phishing websites on the internet, highlighting the growing prevalence of phishing attacks through fraudulent websites. The RSA Quarterly Fraud Report (Q1 2020) also noted an increase in COVID-19-related phishing attacks and virtual entertainment scams worldwide, indicating the significant financial damage and pressure caused by these attacks. While several anti-phishing solutions have been proposed by cybersecurity experts and researchers, blacklist-based identification remains a popular approach. This involves using a blacklist system on web browsers that compares Uniform Resource Locators (URLs) to designated phishing website URLs to verify their authenticity [5]. However, the limitation of this method is its inability to

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

identify new phishing URLs due to the dynamic nature of cyber-attacks. To address this, machine learning-based solutions have been developed to evaluate the legitimacy of websites based on their features, providing flexibility in identifying new phishing sites [6]. However, the effectiveness of these solutions depends on the chosen machine learning method, and some have been found to have low detection precision and high false positive rates [7]. This may be due to data quality issues, such as class imbalance, which can unfairly impact the performance of machine learning strategies [8].

The proposed system presents a promising approach for detecting web attacks on IoT devices using an ensemble classification approach. The authors' experimental results show that their system achieves high accuracy and low false positive rates, making it suitable for deployment in real-world IoT environments [9].

Li et. al [10] propose an edge computing-based approach to implement their system. The IoT devices collect and preprocess data locally, while the anomaly detection algorithms are deployed on edge devices such as gateways or fog nodes. This approach reduces the latency and bandwidth requirements of transmitting data to a centralized server for analysis.

Zhang et. al [11] proposes a location privacy protection scheme based on differential privacy and game theory that can effectively protect users' location privacy in a mobile edge computing environment. The proposed scheme provides a promising approach to address the location privacy challenge in mobile edge computing.

Munezero et. al. [12] present a thorough survey of data mining and machine learning techniques utilized for traffic classification in sustainable smart cities. The authors provide a comprehensive overview of the challenges and limitations associated with implementing these techniques in the context of smart cities. They emphasize the significance of real-time processing of traffic data for effective traffic management. This paper serves as a valuable resource for researchers and practitioners engaged in the development of sustainable smart cities.

This study introduces a meta-learning approach based on ensemble models for detecting phishing websites. The proposed model involves multiple machine learning algorithms trained to combine predictions from other machine learning algorithms, a technique commonly used in the field of ensemble learning to optimize performance [13]. This article makes significant contributions to the existing knowledge by:

- Implementing a range of base learners and their variations to accurately identify both legitimate and phishing websites.
- Developing ensemble-based meta-learners through stacking techniques to further enhance the model's performance.
- Conducting a comprehensive and accurate comparison of the proposed methods with existing state-of-the-art phishing detection techniques, providing valuable insights into the effectiveness of different approaches.

This article aims to answer the following research questions:

- What is the effectiveness of using base learning algorithms to accurately identify phishing and legitimate websites?
- How efficient are meta-learners in identifying both legitimate and fraudulent websites?
- How does the performance of the proposed ensemble-based meta-learners through stacking compare to existing state-of-the-art phishing detection methods?

This article's structure is presented in the following section. In Section 2, related works are discussed. Section 3 provides details on the research methodology, including the investigation theory, the diagram of the proposed models, and the actual computations employed. Section 4 presents the results of the investigation test and evaluations of the exploratory outcomes. In Section 5, the discussion is presented in light of the previous section's results. Finally, Section 6 concludes the article and offers insights into future work.

II. BACKGROUND

In this section, various ML-based phishing revelation techniques are investigated and analyzed.

Aburrous and Khelifi [14] proposed a novel approach for identifying phishing websites using a toolbar equipped with clever fuzzy classification mining techniques. In their study, they used the Phish Tank dataset to evaluate the performance of their identification module, comparing it with other popular anti-phishing toolbars. The results showed that their approach was able to identify approximately 86% of all analyzed phishing websites with a low false positive rate and a reasonable miss rate. This suggests that their method is effective in improving the accuracy of phishing detection and can potentially enhance the security of online users.

Mohammed et al. [15] proposed a new technique for detecting phishing websites. Their model utilizes a general measure that modifies the connectivity structure by adjusting the learning rate before introducing new neurons. The

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

experimental outcomes demonstrated that the model achieved remarkable accuracy rates of 94.07%, 92.48%, and 91.12% on the training, testing, and validation sets, respectively. These results indicate that the proposed approach is highly successful in identifying phishing websites with a considerable degree of accuracy.

Abdul Hamid et al. [16] introduced a novel approach named MCAC-based multi-label classifier and compared its performance with five other popular systems. The study demonstrated that the proposed method outperforms CBA, MCAR, C4.5, PART, and RIPPER by 2.56%, 0.8%, 1.24%, 4.46%, and 1.86% in terms of accuracy, respectively. Additionally, the study showed that MCAC provides higher precision and generates new rules for improving its visual displays.

Verma and Das [17] conducted an analysis in which they employed a Deep Belief Network (DBN) to identify phishing sites. By utilizing Restricted Boltzmann Machines (RBM) to train their models, DBN models are able to generate highly distinctive feature representations from a given dataset. With an accuracy of 94.43 percent, the proposed DBN outperforms Decision Tree (DT) and Random Forest (RF) algorithms in identifying phishing sites.

Shinde et al. [18] proposed a method that combines fuzzy logic with the RIPPER data mining algorithm to detect phishing websites. The Phish Tank dataset, consisting of 100 websites, was used for experimentation. The results showed that the proposed method generated 12 rules and accurately classified about 85.4% of the phishing emails.

Ali and Ahmed [19] conducted a study on identifying phishing websites using a deep neural network (DNN) with features selected by a genetic algorithm (GA). Their approach outperformed several benchmark classifiers, including decision trees (DT), k-nearest neighbors (KNN), support vector machines (SVM), back-induced backpropagation (Back-Induced BP), and naive Bayes (NB). The results of their research demonstrate the effectiveness of the proposed approach for phishing detection.

Hadi and Alwedyan [20] tested various data mining description computations. The results of the tests show that the MCAR calculation is more accurate than any overestimation. In particular, MCAR outperforms NB, SVM, and CBA by 5.4%, 6.8%, and 6.1%, respectively.

Vrbančić et al. [21] proposed a novel approach for DNN development that utilizes the bat meta-heuristic algorithm. The proposed approach achieved a maximum accuracy of 96.99% for identifying phishing websites. The study demonstrates the effectiveness of the proposed approach in improving the accuracy of DNN models for phishing detection.

Aydin and Baykal [22] evaluated the performance of two classification algorithms, Naive Bayes (NB) and Support Vector Machine with a radial basis function kernel (SMO), using two different feature subset selection techniques: the Correlation-based Feature Selection (CFS) and Consistency-based Subset Evaluation (CS). The results indicated that SMO achieved better performance with the Consistency subset procedure, while NB performed better with CFS. Nevertheless, when examining the overall performance, SMO outperformed NB using both feature selection techniques.

Alqahtani [23] introduced a novel association rule-based approach for detecting phishing websites. This method employs an association rule technique to evaluate the legitimacy of a given site. The experimental findings demonstrate the effectiveness of the proposed approach, which outperforms established classifiers, such as decision trees, RIPPER, and some association learning-based representation models. The proposed method achieves an accuracy of 95.20% and an F-measure of 0.9511, highlighting its superior performance.

Vaithiyanathan et al. [24] conducted an investigation on four specific classification techniques - J48, MLP, NB-updatable, and Bayesian networks using three modified datasets: labor, soybean, and environment. The authors found that on the environment and soy datasets, the Naive Bayes updateable outperformed, while on the labor dataset, MLP performed well in terms of accuracy. Both J48 and NB were found to be highly efficient. The study also highlighted that the performance of different classifiers varies depending on the heuristic list, and the factors that influence the performance of specific classifiers were elaborated upon, including guiding classification, quality level, structure, and type of features.

Pandey et. al. [25] presents a machine learning approach for predicting phishing websites. The authors collected a dataset of 8,000 URLs, half of which were phishing and the other half were legitimate. They used various features related to the URLs, such as domain length, presence of certain characters, and use of subdomains, to train and test six different machine learning models: Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbors, Naive Bayes, and Support Vector Machine. The Random Forest model performed the best with an accuracy of 97.9%. The study shows that machine learning can be an effective method for predicting phishing websites and could be used to improve online security.

Abu-nimeh et al. [26] evaluated the performance of six classifiers (LR, CART, BART, SVM, RF, and NNet) on a dataset of 43 features for phishing detection. The final result

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

showed that RF performed the best with an error rate of 7.72%, although it had the highest spoofing positive rate at 8.29%. LR performed well with a weighted error rate of 3.82% when utilizing cost-sensitive metrics. The authors also evaluated the classifiers using the "area under the curve" metric and found that it was a universal necessity for all classifiers.

Dedakia and Mistry [27]. By taking into account realized features, the proposed method improves upon the MCAC method. The experimental results indicate that the proposed CBAC system is accurate to 94.29 percent.

Wedyan and Wedyan [28] proposed the Phishing Familiarity Program as a related phishing detection algorithm. They evaluated the performance of the proposed algorithm against four well-known classifiers: C4.5, PRISM, CBA, and MCAR, using 17 features. The results show that the PAC classifier outperformed the others in both rare and frequent feature sets, with an accuracy of 99.31%. The accuracy of PAC was comparable to that of MCAR, but the proposed PAC algorithm provided a higher accuracy rate.

Abikoye et al. [29] presents a modified version of the Advanced Encryption Standard (AES) algorithm for information security. The authors propose modifications to the S-box and key schedule of the AES algorithm to enhance its resistance against attacks such as brute force and differential attacks. The performance of the modified AES algorithm is evaluated using various metrics such as encryption speed, avalanche effect, and key sensitivity. The experimental results show that the proposed modifications improve the security and efficiency of the AES algorithm.

Anh et al. [30] proposes a fuzzy-based system for identifying phishing attacks using a dataset of 11,660 instances as the test reasons and ten datasets with 1,000 instances of phishing and 1,000 instances of benign objects as the training reasons. The proposed method achieves an accuracy of 99.25%. The authors provide a detailed explanation of the fuzzy-based system and its implementation for phishing identification. They also compare their proposed method with other existing techniques and discuss the advantages and limitations of their approach. Overall, the paper presents an efficient and effective approach to identifying phishing attacks using fuzzy logic.

The paper by Rahman et al. [31] examines the performance of various machine learning classifiers for detecting phishing URLs. Specifically, they analyze the effectiveness of K-NN, decision trees (DT), support vector machines (SVM), random forest (RF), extremely randomized trees (ERT), and gradient boosting trees (GBT) in identifying phishing URLs. The authors conducted experiments using a dataset of 11,055 instances, and their results showed that RF achieved the highest accuracy of 98.26%, followed by SVM with an accuracy of 97.69%. The study concludes that ML techniques can effectively detect phishing URLs and recommends the use of RF and SVM for practical applications.

Aberus et al. [32] proposed a theory for phishing website detection using six classes of 20 selected features. They tested the theory using six classifiers: JRip, Part, PRISM, C4.5, CBA, and MCAR. The experimental results show that CBA and MCAR outperform the standard classifiers in terms of accuracy. MCAR with 22 rules was found to be the most accurate classifier, outperforming any other conventional classifier.

III. METHODOLOGY

The method of investigation used in this research paper is described in this section. Particularly, the examined exploratory structure, phishing datasets, proposed approaches, and assessment measurements.

A. Baseline Classifiers

In the field of machine learning, there are various algorithms available to choose from, each with different levels of accuracy and effectiveness [33]. To assess the performance of other machine learning algorithms, a benchmark prediction algorithm provides a set of predictions that can be used for comparison. In this study, several classifiers including Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbor (K-NN), Decision Tree (DT), Support Vector Machine (SVM), Bagging, and Adaboost were utilized [34]. The main aim of this study was to evaluate the performance of these classifiers for predicting phishing websites.

B. Ensemble Classifiers

Bagging: This article utilizes the bagging strategy as a meta-learner to enhance the performance of prediction models [35]. Bagging involves training prediction models using experiences derived from various subsets of the original dataset. This technique aims to reduce the variation in the generated models while preventing an increase in bias by applying an accumulation strategy on all the models produced. Additionally, bagging enhances performance by randomly re-sampling the dataset and generating different base models by fitting them on the re-sampled subsets. Finally, bagging aggregates the base models into a single prediction process [36]. Thus, the Bagging algorithm is employed in this study.

Bagging Algorithm

Inputs:

X: training data features

y: training data labels

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
Article Received: 26 June 2023 Revised: 22 August 2023 Accepted: 08 September 2023

M: number of models to train

model_algorithm: base machine learning algorithm to use

Outputs:

List of trained models

Initialize an empty list of models.

For m = 1 to M:

Create a bootstrap sample of X and y by randomly selecting N samples with replacement from the original dataset, where N is the size of the dataset.

Train a base model on the bootstrap sample using the specified model algorithm.

Add the trained model to the list of models.

Return the list of trained models.

AdaBoost: Adaptive Boosting (AdaBoost) is a meta-learning algorithm that applies a weak standard learner sequentially to a reweighted dataset. The algorithm works by creating a decision tree and then training subsequent trees based on the error of the previous trees. The amount of attention paid to each subsequent tree is determined by the error of the previous trees. This process ensures that data that is difficult to predict receives greater weight, while data that is easy to predict receives less weight [37-38].

AdaBoost is specifically designed for binary classification purposes, making it a suitable algorithm for detecting phishing websites. By using this algorithm, we can improve the accuracy of our classification model, making it more effective at identifying phishing websites.

Inputs:

X: training data features

y: training data labels

M: number of base models to train

base_algorithm: the base machine learning algorithm to use

Outputs:

List of trained models

List of corresponding model weights

Initialize the weights for each training example to 1/n, where n is the number of training examples.

For m = 1 to M:

Train a base model on the training data using the specified base algorithm, weighted by the sample weights from the previous iteration.

Calculate the error of the base model on the training data by summing the weights of misclassified examples.

Calculate the weight of the current base model using the formula: $\alpha(m) = 0.5 * \log\frac{(1-\text{error}(m))}{\text{error}(m)}$

Update the weights of the training examples using the formula:

$w(i) = w(i)\exp\frac{(-\alpha(m)*y(i)*h(m)(x(i))}{z}$ where Z is a normalization constant

Add the current model and its corresponding weight $\alpha(m)$ to the list of models and weights.

Return the list of trained models and corresponding weights.

C. Combination of Baseliner and Ensemble Classifiers

To ensure high accuracy and low error rates, we evaluate the performance of classifiers individually. We then combine the classifiers based on their performance to create an optimized model (meta-learner) [39]. The combination of classifiers used in this article is (LR + Bagging, RF + Bagging, RF + K-NN, K-NN + DT, DT + SVM, SVM + AdaBoost).

By combining these classifiers, we can improve the accuracy and reliability of our model. This approach allows us to leverage the strengths of each individual classifier, while mitigating their weaknesses. Overall, the resulting optimized model is better equipped to accurately classify phishing websites, providing enhanced protection against cyber threats.

D. Stacking of Classifiers

To enhance the efficacy of our classifiers, we employ a technique known as stacked or stack ensemble learning, which is a type of machine learning algorithm [40]. This method involves utilizing a meta-learning algorithm to identify the most effective method for combining predictions from two or more base machine learning algorithms.

The main advantage of stacking is that it enables us to leverage the strengths of multiple well-performing models in an ensemble. By doing so, we can create predictions that outperform any single model in the ensemble, resulting in better overall performance. This technique is particularly useful for classification or regression tasks, as it allows us to harness the capabilities of multiple models to improve the accuracy and reliability of our predictions.

E. Experiment Structure

This section outlines the preliminary procedure for the inspection, as shown in Fig 1. The goal is to observe and test

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

the proposed phishing site validation procedures, and the exploratory system was coordinated to achieve this.

To design and evaluate the proposed system, a phishing dataset from the UCI repository was used. K-fold cross-validation methods were employed to create and evaluate phishing models. The ability of the 10-fold CV decision to minimize the impact of the class imbalance issue was critical, and the K-fold CV strategy ensured that each model could be used iteratively for training and testing [41].

The phishing dataset was subjected to 10-fold CV, followed by the proposed technique and the selected metric classifier. The effectiveness of the created phishing model's phishing recognition was then tested to differentiate it from other well-established phishing recognition methods. All experiments were carried out using an AI tool developed in Python 3.7, in the same environment to ensure consistency.

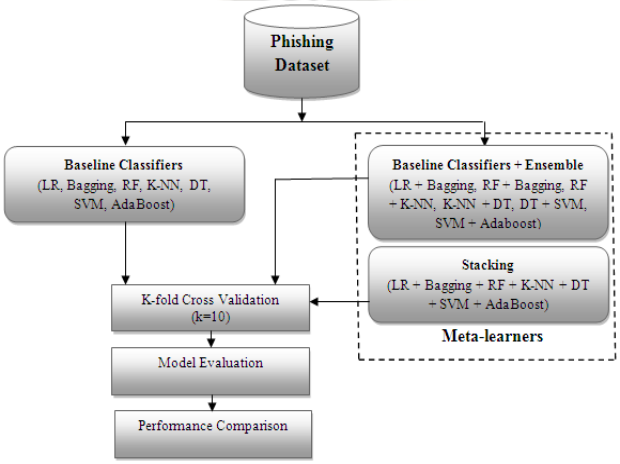

Figure 1. Experimental Structure of Methodology

F. Phishing Datset

During the training and testing phase of this review, a phishing dataset was used, which is readily available and commonly used in current studies [42]. The dataset contains 11,054 instances, including 4,897 phishing and 6,157 legitimate ones. The dataset contains 30 different features that characterize it.

G. Performance of Evaluation Metrics

To evaluate the performance of the created phishing models, their accuracy, precision, recall, f1-score, confusion matrix, and receiver operator characteristic (ROC) curve are utilized as evaluation metrics [43]. The choice of these metrics is based on their wide and common usage for evaluating phishing site detection in existing studies.

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

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

$$\text{F1 score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

The ROC curve is a graphical representation of how the true positive rate (TPR) and false positive rate (FPR) change as the classification threshold is varied. An ideal classifier would have a ROC curve passing through the top left corner of the plot, where TPR equals 1 and FPR equals 0. The area under the ROC curve (AUC) provides a single metric to summarize the classifier's overall performance, where a perfect classifier has an AUC of 1, while a random classifier has an AUC of 0.5.

IV. RESULTS

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

This section presents the results of our experiments aimed at answering the research questions. Table 1 presents the experimental results, where P represents Phishing and N represents Non-Phishing. As demonstrated in Table 1, the bagging classifier performed better than the other classifiers in terms of accuracy and other performance evaluation metrics.

TABLE I. PERFORMANCE OF DIFFERENT BASE LINER CLASSIFIERS WITH THEIR CORRESPONDING METRICS

Model	% Accuracy	P / N	Precision	Recall	f1-score	Confusion Matrix [[TP FP] [FN TN]]
LR	92.67	0	0.92	0.91	0.91	[[856 86] [76 1193]]
		1	0.93	0.94	0.94	
Bagging	97.15	0	0.97	0.96	0.97	[[907 35] [28 1241]]
		1	0.97	0.98	0.98	
RF	92.53	0	0.93	0.89	0.91	[[838 104] [61 1208]]
		1	0.92	0.95	0.94	
K-NN	95.20	0	0.95	0.94	0.94	[[884 58]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
Article Received: *26 June 2023* ***Revised:*** *22 August 2023* ***Accepted:*** *08 September 2023*

Model	% Accuracy	P / N	Precision	Recall	f1-score	Confusion Matrix [[TP FP] [FN TN]]
		1	0.95	0.96	0.96	[48 1221]]
DT	92.26	0	0.89	0.94	0.91	[[881 61]
		1	0.95	0.91	0.93	[110 1159]]
SVM	95.97	0	0.96	0.94	0.95	[[886 56]
		1	0.96	0.97	0.97	[33 1236]]
Adaboost	94.02	0	0.94	0.92	0.93	[[870 72]
		1	0.94	0.95	0.95	[60 1209]]

Figure 2 and Figure 3 also represent the classifiers efficiency as well as precision, recall and f1-measure for better understanding of the performance.

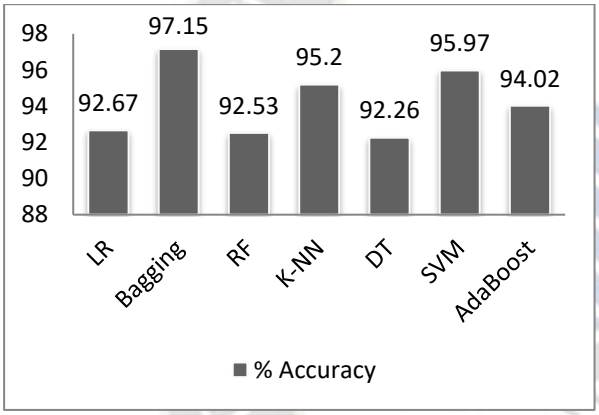

Figure 2. Accuracy Presentation of Multiple Classifiers

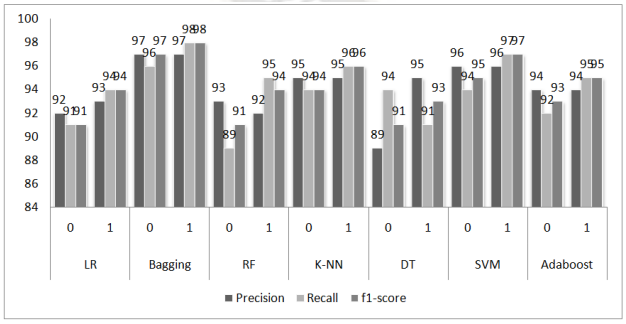

Figure 3. Precision, Recall and f1-score of Multiple Classifiers

In Figure 4, a ROC (AUC) curve has been drawn between obvious positive rate and false positive rate. In the series of the presentation of the individual classifiers bagging has the highest ROC (AUC) value 99.2%.

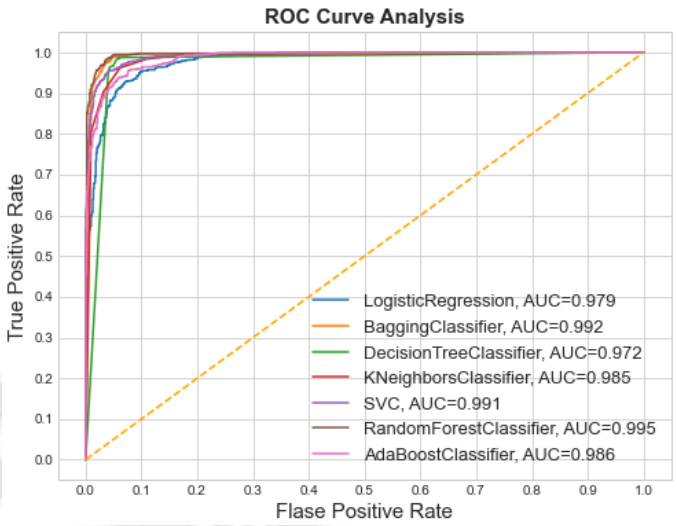

Figure 4. ROC Curve (AUC) in between Multiple Classifiers

As shown in Table 2, the performance of the proposed model and its corresponding evaluation metrics are presented. The proposed model utilizes a combination of various classifiers and stacks them using ensemble learning. The accuracy and other relevant performance metrics of the proposed model are illustrated in Figure 5 and Figure 6, respectively.

TABLE II. PERFORMANCE OF PROPOSED MODEL AND THEIR METRICS

Proposed Model	% Accuracy	P / N	Precision	Recall	f1-score	Confusion Matrix [[TP FP] [FN TN]]
LR + Bagging	96.78	0	0.96	0.96	0.96	[[904 38]
		1	0.97	0.97	0.97	[33 1236]]
Bagging + RF	97.01	0	0.97	0.96	0.96	[[908 34]
		1	0.97	0.97	0.97	[32 1237]]
RF + K-NN	94.16	0	0.91	0.96	0.93	[[907 35]
		1	0.97	0.93	0.95	[94 1175]]
K-NN + DT	94.30	0	0.97	0.89	0.93	[[839 103]
		1	0.92	0.98	0.95	[23 1246]]
DT + SVM	95.97	0	0.96	0.94	0.95	[[886 56]
		1	0.96	0.97	0.97	[33 1236]]
SVM + Adaboost	95.97	0	0.96	0.94	0.95	[[886 56]
		1	0.96	0.97	0.97	[33 1236]]
LR + Bagging + RF +K-NN + DT + SVM + Adaboost	97.19	0	0.97	0.96	0.97	[[906 36]
		1	0.97	0.98	0.98	[26 1243]]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

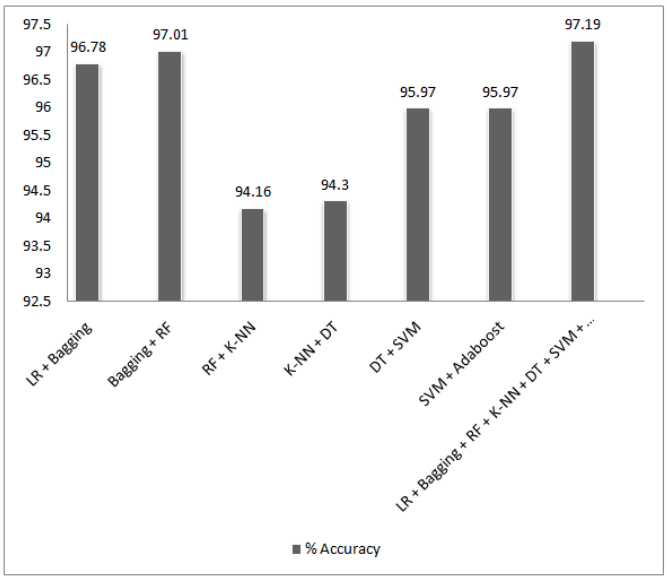

Figure 5. Accuracy of Proposed Model

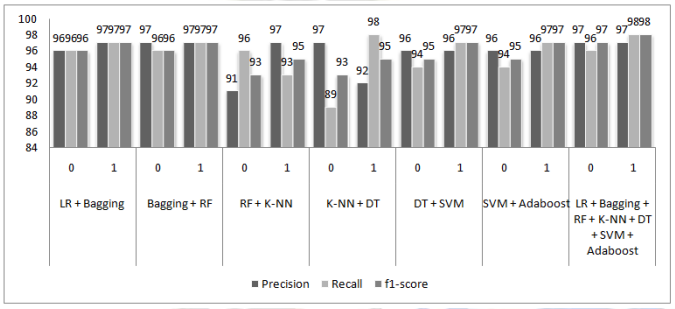

Figure 6. Precision, Recall and f1-score of Proposed Model

V. DISCUSSION

In this study, we conducted experiments to evaluate the effectiveness of various ensemble meta-learners for stacking in the proposed phishing identification model. We compared the performance of our model with that of classifiers from previous studies, using the dataset provided in this article. Table 1 shows the results of the baseline classifiers, while Table 2 presents the results of the proposed model (LR + Bagging, Bagging + RF, RF + K-NN, K-NN + DT, DT + SVM, SVM + Adaboost) and the stacking model (LR + Bagging + RF + K-NN + DT + SVM + Adaboost) in comparison with benchmark classifiers. We found that the proposed model outperformed the standard classifiers in terms of accuracy and other performance metrics.

Table 1 shows the accuracy and other performance metrics of the baseline classifiers, including LR, Bagging, RF, K-NN, DT, SVM, and Adaboost. Among these classifiers, Bagging achieved the best accuracy and other metrics, with an accuracy of 97.15 and precision, recall, and F1-score values of 0.93, 0.89, and 0.91, respectively. Fig 6 illustrates the ROC (AUC) curve for the Bagging classifier, which is the second best after the RF classifier (99.5) in terms of AUC value. Thus, Bagging can be considered as the best classifier for phishing identification among the individual classifiers.

Table 2 presents the results of the proposed model and various combinations of ensemble meta-learners. The combination of Bagging and RF achieved the highest accuracy (97.01) among all the combinations. This combination also yielded the highest precision, recall, and F1-score values compared to other meta-learners. Therefore, the combination of Bagging and RF can be used as a meta-learner for phishing identification.

Finally, we applied the ensemble of meta-learners for stacking to the dataset for phishing detection. The stacking model achieved a higher accuracy (97.19) and other performance metrics than the baseline classifiers and the proposed model. All experiments were conducted under the same conditions, such as 10-fold cross-validation and the same number of instances for training and testing datasets. Fig 5 and Fig 6 show the accuracy and performance metrics of the different models, respectively.

These results suggest that the ensemble meta-learners for stacking and meta-learners in the proposed model are superior to the baseline models in terms of accuracy and other performance metrics. Our stacking model can effectively distinguish phishing sites from legitimate sites with a high degree of accuracy and a reduced error rate.

In summary, to answer the research objectives, we found that the base learning algorithms (RG1) performed well in identifying phishing and legitimate sites. The meta-learners (RG2) significantly improved the performance of the classifiers, particularly when Bagging was combined with the RF classifier. Our proposed model and stacking model outperformed existing state-of-the-art methods (RG3) in phishing detection.

VI. CONCLUSION

This study aims to investigate the effectiveness of an ensemble of meta-learners for stacking in the context of phishing site detection. Three types of models, as outlined in Section 3, were tested and analyzed for their ability to identify phishing sites. The experimental results indicate that the proposed ensemble-based meta-learners for stacking outperformed existing phishing site detection models. This approach showcases the potential of meta-learners as an intelligent algorithm for designing models with greater accuracy and reliability in distinguishing phishing sites. The proposed technique achieved an impressive predictive accuracy of approximately 97.19 percent, as well as high precision, recall, and f1-score. These findings demonstrate the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

effectiveness and reliability of the proposed approach in maintaining high detection accuracy.

To further validate the proposed method, future work will apply it to a consistent dataset to mitigate the impact of phishing websites on the internet. Additionally, research will explore how the quality of the dataset, class inconsistencies, and high-dimensional issues may affect the detection of phishing sites in different regions.

REFERENCES

[1] Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. Applied Geography, 49, 45-53.

[2] Thabit, F., Alhomdy, S. A. H., Alahdal, A., & Jagtap, S. B. (2020). Exploration of Security Challenges in Cloud Computing: Issues, Threats, and Attacks with their Alleviating Techniques. Journal of Information and Computational Science, 12(10).

[3] Auerbach, S. (2008). Screening out cyberbullies: Remedies for victims on the internet playground. Cardozo L. Rev., 30, 1641.

[4] Karsten, P., & Bateman, O. (2016). Detecting Good Public Policy Rationales for the American Rule: A Response to the Ill-Conceived Calls for Loser Pays Rules. Duke LJ, 66, 729.

[5] Sountharrajan, S., Nivashini, M., Shandilya, S. K., Suganya, E., Bazila Banu, A., & Karthiga, M. (2020). Dynamic recognition of phishing URLs using deep learning techniques. In Advances in cyber security analytics and decision systems (pp. 27-56). Springer, Cham.

[6] Mourtaji, Y., Bouhorma, M., Alghazzawi, D., Aldabbagh, G., & Alghamdi, A. (2021). Hybrid rule-based solution for phishing URL detection using convolutional neural network. Wireless Communications and Mobile Computing, 2021.

[7] Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2021). Phishing email detection using natural language processing techniques: a literature survey. Procedia Computer Science, 189, 19-28.

[8] Ardalani, H., Vidkjær, N. H., Kryger, P., Fiehn, O., & Fomsgaard, I. S. (2021). Metabolomics unveils the influence of dietary phytochemicals on residual pesticide concentrations in honey bees. Environment International, 152, 106503.

[9] Liu, C., Wei, H., Qiu, T., & Zhu, X. (2018). A novel web attack detection system for Internet of Things via ensemble classification. IEEE Access, 6, 64594-64606.

[10] Li, Y., Zhang, S., Chen, Y., & Chen, J. (2021). An edge computing based anomaly detection method in IoT industrial sustainability. IEEE Transactions on Industrial Informatics, 17(3), 2053-2062.

[11] Wang, L., Wang, C., Cai, Z., Zhang, J., & Chen, W. (2019). Location privacy challenge in mobile edge computing. IEEE Network, 33(6), 52-58.

[12] Munezero, M., Crespi, N., & Zeadally, S. (2020). Data mining and machine learning methods for sustainable smart cities traffic classification: A survey. Sustainable Cities and Society, 53, 101973.

[13] Antonopoulos, I., Robu, V., Couraud, B., Kirli, D., Norbu, S., Kiprakis, A., ... & Wattam, S. (2020). Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. Renewable and Sustainable Energy Reviews, 130, 109899.

[14] Aburrous, M., & Khelifi, A. (2013, March). Phishing detection plug-in toolbar using intelligent Fuzzy-classification mining techniques. In The international conference on soft computing and software engineering [SCSE'13], San Francisco State University, San Francisco, California, USA.

[15] Prabha , G. ., Mohan, A. ., Kumar, R. D. ., & Velrajkumar, G. . (2023). Computational Analogies of Polyvinyl Alcohol Fibres Processed Intellgent Systems with Ferrocement Slabs. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 313–321. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2669.

[16] Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications, 25(2), 443-458. doi: 10.1007/s00521-013-1491-8

[17] Abdelhamid, N., Ayesh, A., & Thabtah, F. (2014). Phishing detection based associative classification data mining. Expert Systems with Applications, 41(13), 5948-5959.

[18] Verma, R., & Das, A. (2017, March). What's in a url: Fast feature extraction and malicious url detection. In Proceedings of the 3rd ACM on International Workshop on Security and Privacy Analytics (IWSPA) (pp. 55-63). ACM.

[19] Khadi, A., & Shinde, S. (2014). Detection of phishing websites using data mining techniques. International Journal of Engineering Research and Technology, 2(12), 3725-3729.

[20] Ali, W., & Ahmed, A. A. (2019). Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting. IET Information Security, 13(6), 659-669.

[21] Moh'd Iqbal, A. L., Hadi, W. E., & Alwedyan, J. (2013). Detecting Phishing Websites Using Associative Classification. Journal of Information Engineering and Applications, VoI, 3.

[22] Vrbančič, G., Fister Jr, I., & Podgorelec, V. (2018, June). Swarm intelligence approaches for parameter setting of deep learning neural network: case study on phishing websites classification. In Proceedings of the 8th international conference on web intelligence, mining and semantics (pp. 1-8).

[23] Aydin, M., & Baykal, N. (2015, September). Feature extraction and classification phishing websites based on URL. In 2015 IEEE Conference on Communications and Network Security (CNS) (pp. 769-770). IEEE.

[24] Alqahtani, M. (2019, April). Phishing websites classification using association classification (PWCAC). In 2019 International conference on computer and information sciences (ICCIS) (pp. 1-6). IEEE.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8164
***Article Received:** 26 June 2023 **Revised:** 22 August 2023 **Accepted:** 08 September 2023*

[25] Vaithiyanathan, V., Rajeswari, K., Tajane, K., & Pitale, R. (2013). Comparison of different classification techniques using different datasets. International Journal of Advances in Engineering & Technology, 6(2), 764.

[26] Pandey, M. K., Singh, M. K., Pal, S., & Tiwari, B. B. (2023). Prediction of phishing websites using machine learning. Spatial Information Research, 31(2), 157-166.

[27] Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007, October). A comparison of machine learning techniques for phishing detection. In Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit (pp. 60-69).

[28] Dedakia, M., & Mistry, K. (2015). Phishing detection using content based associative classification data mining. J. Eng. Comput. Appl. Sci, 4(7), 209-214.

[29] Wedyan, S., & Wedyan, F. (2013). An Associative Classification Data Mining Approach for Detecting Phishing Websites. Journal of Emerging Trends in Computing and Information Sciences, 4(12).

[30] Abikoye, O. C., Haruna, A. D., Abubakar, A., Akande, N. O., & Asani, E. O. (2019). Modified advanced encryption standard algorithm for information security. Symmetry, 11(12), 1484.

[31] Nguyen, L. A. T., & Nguyen, H. K. (2015, May). Developing an efficient fuzzy model for phishing identification. In 2015 10th Asian Control Conference (ASCC) (pp. 1-6). IEEE.

[32] Rahman, S. S. M. M., Rafiq, F. B., Toma, T. R., Hossain, S. S., & Biplob, K. B. B. (2020). Performance assessment of multiple machine learning classifiers for detecting the phishing URLs. In Data Engineering and Communication Technology: ICDECT 2019 (L. B. Das, S. Mukhopadhyay, & V. K. Singh, Eds.) (pp. 285-296). Springer.

[33] Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010, April). Predicting phishing websites using classification mining techniques with experimental case studies. In 2010 Seventh International Conference on Information Technology: New Generations (pp. 176-181). IEEE.

[34] Law, E., & Ahn, L. V. (2011). Human computation. Synthesis lectures on artificial intelligence and machine learning, 5(3), 1-121.

[35] Chaurasia, V., & Pal, S. (2020). Machine learning algorithms using binary classification and multi model ensemble techniques for skin diseases prediction. International Journal of Biomedical Engineering and Technology, 34(1), 57-74.

[36] Livieris, I. E., Pintelas, E., Stavroyiannis, S., & Pintelas, P. (2020). Ensemble deep learning models for forecasting cryptocurrency time-series. Algorithms, 13(5), 121.

[37] Ahmadi, A., Nabipour, M., Mohammadi-Ivatloo, B., Amani, A. M., Rho, S., & Piran, M. J. (2020). Long-term wind power forecasting using tree-based learning algorithms. IEEE Access, 8, 151511-151522.

[38] Chen, C. H., Tanaka, K., Kotera, M., & Funatsu, K. (2020). Comparison and improvement of the predictability and interpretability with ensemble learning models in QSPR applications. Journal of cheminformatics, 12(1), 1-16.

[39] Sneha, N., & Gangil, T. (2019). Analysis of diabetes mellitus for early prediction using optimal features selection. Journal of Big data, 6(1), 1-19.

[40] Alejandro Garcia, Machine Learning for Customer Segmentation and Targeted Marketing , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[41] Wang, Y. X., Girshick, R., Hebert, M., & Hariharan, B. (2018). Low-shot learning from imaginary data. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7278-7286).

[42] Chaurasia, V., Pandey, M. K., & Pal, S. (2022). Chronic kidney disease: a prediction and comparison of ensemble and basic classifiers performance. Human-Intelligent Systems Integration, 1-10.

[43] Chen, K. Y., Marschall, E. A., Sovic, M. G., Fries, A. C., Gibbs, H. L., & Ludsin, S. A. (2018). assign POP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. Methods in Ecology and Evolution, 9(2), 439-446.

[44] Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications, 25(2), 443-458.

[45] Adil, M., Javaid, N., Qasim, U., Ullah, I., Shafiq, M., & Choi, J. G. (2020). LSTM and bat-based RUSBoost approach for electricity theft detection. Applied Sciences, 10(12), 4378.