Rotation Forest-Based Logistic Model Tree for Website Phishing Detection

Abdullateef O. Balogun[1,2], Noah O. Akande[3(✉)], Fatimah E. Usman-Hamza[1], Victor E. Adeyemo[4], Modinat A. Mabayoje[1], and Ahmed O. Ameen[1]

[1] Department of Computer Science, University of Ilorin, Ilorin PMB 1515, Nigeria
{balogun.ao1,usman-hamzah.fe,mabayoje.ma,aminamed}@unilorin.edu.ng
[2] Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar 32610, Perak, Malaysia
abdullateef_16005851@utp.edu.my
[3] Department of Computer Science, Landmark University, Omu-Aran, Kwara State, Nigeria
akande.noah@lmu.edu.ng
[4] School of Built Environment, Engineering, and Computing, Leeds Beckett University, Headingley Campus, Leeds LS6 3QS, UK
v.adeyemo5225@student.leedsbeckett.ac.uk

Abstract. The emergence of web and internet technology has led to its use in a broad array of services ranging from financial to educational services. This has led to a spike in the number of cybersecurity problems over the years, the most notable of which is the phishing attack, in which malicious websites imitate legitimate websites to capture gullible users' details needed for unauthorized access. However, current mitigation strategies, such as anti-phishing applications and Machine Learning (ML) methods, have been effective for detecting phishing activities. Hackers, on the other hand, are developing new ways to circumvent these countermeasures. Nevertheless, given the dynamism of phishing attempts, there is a continual demand for innovative and efficient solutions for website phishing detection. This study proposes a Rotation Forest-based Logistic Model Trees (RF-LMT) for website phishing detection. LMT is a technique that combines logistic regression and tree inference into a single model tree. Three datasets of different instance distributions, both balanced and imbalanced, are used to investigate the proposed RF-LMT. From the results, it was observed that LMT performed better than the selected baseline classifiers. This finding revealed that LMT can perform comparably to baseline classifiers. However, in comparison to LMT and experimented baseline classifiers, the proposed RF-LMT method showed superior performance in website phishing detection. Specifically, RF-LMT had a high detection accuracy (98.24%), AUC (0.998), f-measure (0.982) values with a low false-positive rate (0.018). Furthermore, RF-LMT outperformed existing ML-based phishing attack models. As a result, the proposed RF-LMT method is recommended for dealing with complex phishing attacks.

Keywords: Cybersecurity · Logistic Model Tree · Machine learning · Phishing attack · Rotation forest

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12957, pp. 154–169, 2021.
https://doi.org/10.1007/978-3-030-87013-3_12

1 Introduction

The increased availability and application of Information Technology (IT) have increased the number of internet-based applications available in cyberspace. These operations range from vital services such as financial services to essential activities such as health and education applications [1, 2]. Financial purchases, online gaming platforms, and social media apps, according to data, are among the most popular and commonly used internet-based solutions with a large user base. The vast number of users who use these internet-based solutions demonstrate their recent successes.

According to research, financial transactions, online gaming sites, and social media applications are among the most common and widely used web-based solutions with a broad user base. The large number of people who use these web-based applications demonstrate their popularity in recent years. The aim is to increase the accessibility and availability of commonly used internet-based solutions. Nonetheless, since there are no generic cyberspace control mechanisms, the unrestricted mobility and affordability of these internet-based solutions in cyberspace open the door to cyber-attacks [3–5]. Cyber-attacks generate critical vulnerabilities and risks for both internet-based solutions and end-users, as well as important information and financial losses. Phishing attacks on websites are a typical example of these cyber-attacks. Cybercriminals are now setting up bogus websites to steal personal information from unsuspecting users and use it for illegal purposes [2, 6].

The website phishing attack is a significant cybersecurity issue that has overburdened cyberspace and has harmed internet users and internet-based solutions [7, 8]. According to [2], website phishing is a common deception in which an unauthorized website imitates a legitimate website for the sole intention of collecting data from unsuspecting users. As a result, phishing attacks pose a severe risk to web-based solutions [9–11]. In 2018, the Anti-Phishing Working Group (APWG) identified 51,401 phishing websites in cyberspace. According to RSA, international organizations lose almost $9 billion in 2016 due to phishing attacks [12, 13]. These incidents have shown that phishing attacks from unauthorized websites quickly gain ground, resulting in significant financial losses and burdens [9, 11, 14].

Numerous cybersecurity specialists and analysts have proposed and created various anti-phishing methods for identifying phishing websites [15–17]. One of these solutions is the use of a blacklist technique to avoid website phishing attacks. Web browsers' blacklisting mechanism matches the submitted universal resource locator (URLs) with previously-stored phishing website URLs to determine its authenticity. A significant disadvantage of blacklist anti-phishing methods is their failure to detect new phishing URLs due to their reliance on compiling blacklisted phishing URLs [3, 18]. Furthermore, cyber-attackers are deploying sophisticated techniques that enable them to circumvent the blacklisting process easily. Due to the dynamism of cyber-attacks, Machine Learning (ML)-based technologies are used to assess the credibility of websites to handle the complex existence of website phishing attacks on features derived from websites [12, 15, 19].

On the other hand, the efficacy of the ML-based phishing detection method depends on the success of the selected ML technique when detecting phishing websites. Several ML methods have been used to detect phishing websites, with low detection accuracy

and high false-positive rates [6, 20–22]. This might be attributed to difficulties with data quality, like imbalanced datasets, that degrade the effectiveness of ML models [23, 24]. As a result of the dynamism of phishing websites, more sophisticated ML methods are needed.

Consequently, a rotation forest-based logistic model tree (RF-LMT) for identifying phishing websites is proposed. LMT is a model tree that integrates logistic regression and tree induction approaches. The cornerstone of LMT is the incorporation of a logistic regression model at the leaf nodes of the tree by systematically optimizing higher leaf nodes.

Summarily, the following are the specific contributions of this study:

1) RF-LMT algorithm is used to distinguish between legitimate and phishing websites.
2) An experimental evaluation and analysis of RF-LMT for website phishing detection in comparison to existing phishing approaches.

Furthermore, this research aims to address the following research questions:

1) How efficient is the LMT algorithm in detecting legitimate and phishing websites?
2) How efficient is the proposed RF-LMT algorithm in detecting legitimate and phishing websites?
3) How efficient is the proposed RF-LMT compared to existing phishing methods?

The rest of this paper is structured as follows. Section 2 examines existing related research. Section 3 portrays the analysis methodology, an overview of the experimental process, and the algorithms deployed. Section 4 discusses the research experiment and the analysis of the experimental findings. Finally, Sect. 5 concludes and suggests potential future works.

2 Related Works

This section investigates and discusses emerging phishing detection methods developed using different anti-phishing and ML techniques.

Mohammad, Thabtah and McCluskey [1] used a self-structuring neural network to identify phishing websites. Their model is based on an adaptive learning rate that varies before introducing new neurons and network structures. The suggested model's accuracy values were 94.07%, 92.48%, and 91.12% for the training, testing, and validation sets, respectively. Also, the bat meta-heuristics search algorithm was used by Vrbančič, Fister Jr and Podgorelec [2] to boost DNN. The proposed method had a maximum accuracy of 96.9%. These studies demonstrate that neural network models are almost as good as standard classifiers at detecting phishing websites.

Alqahtani [6] identified phishing websites using a novel association law induction strategy. The proposed solution employs an association law procedure to determine the authenticity of a page. Their experimental results showed the effectiveness of the proposed approach, as it outperforms baseline classifiers including DT, RIPPER, and some associative learning classification models with a precision of 95.20% and an F-measure

value of 0.9511. Similarly, Abdelhamid, Ayesh and Thabtah [7] used a Multi-label Classifier-based Associative Classification (MCAC) technique to identify phishing. The MCAC technique was used for the detection mission to remove sixteen (16) unique features from a website URL using rules discovery, classifier creation, and class assignment. From their experimental results, MCAC outperformed the base classifiers RIPPER, DT, Component, CBA, and MCAR. Dedakia and Mistry [8] proposed a Content-Based Associative Classification (CBAC) approach for detecting phishing. The proposed method extends the Multi-Label Class Associative Classification (MCAC) algorithm by considering content-based properties. Based on the experimental results, the proposed solution (CBAC) had an accuracy value of 94.29%. Hadi, Aburub and Alhawari [10] created and tested a fast associative classification algorithm (FACA) for phishing website recognition against other known associative classification (AC) methods (CBA, CMAR, MCAR, and ECAR). Their experimental results show that FACA outperforms other AC methods in terms of accuracy and F-measure values. The effectiveness of these associative-based approaches shows their applicability for phishing detection. However, their low accuracy value is a disadvantage, and high detection accuracy phishing detection models are needed.

Rahman, Rafiq, Toma, Hossain and Biplob [11] investigated the effectiveness of various ML methods and ensemble methods in detecting website phishing (KNN, DT, SVM, RF, Extreme Randomized Tree (ERT), and Gradient Boosting Tree (GBT)). Similarly, Chandra and Jana [9] explored the usage of meta-classifiers to improve the detection of phishing websites. Their analyses showed that ensemble methods outperformed single classifiers. Alsariera, Elijah and Balogun [12] developed ensemble variants of Forest Penalizing by Attributes (ForestPA) to detect phishing websites. Forest employs weight assignment and an increment technique to grow healthy trees. According to their results, the proposed meta-learner ForestPA variants are very good at detecting phishing websites, with a minimum accuracy of 96.26%. Chiew, Tan, Wong, Yong and Tiong [13] proposed a Hybrid Ensemble FS (HEFS) approach based on a novel cumulative distribution function gradient (CDF-g) method to choose optimal functions. The RF estimation of HEFS was 94.6% accurate. Aydin and Baykal [14] used subset-based functionality extracted from a website URL to detect phishing. The extracted features were analyzed using alpha-numeric character, keyword, security, domain identity, and rank-based methods. The extracted features were then subjected to NB and Sequential Minimal Optimization (SMO). Precision was 83.96% for NB and 95.39% for SMO, respectively.

Ubing, Jasmi, Abdullah, Jhanjhi and Supramaniam [17] proposed a phishing approach focused on feature selection (FS) and Ensemble Learning Mechanism (ELM). The Random Forest Regressor (RFG) was used as the FS method, and the ELM was determined by majority voting. Their experimental findings revealed that the proposed methods outperform and perform comparably to existing baseline and ensemble methods.

As a result of the foregoing analyses, there is a need for more reliable and efficient solutions, as the majority of present approaches are relatively ineffective. Therefore, an RF-LMT method is proposed in this study for detecting phishing websites.

3 Methodology

This section describes the experimental methodology used in this study—specifically, Logistic Model Tree (LMT) and the proposed RF-LMT website phishing detection technique. The phishing datasets used for training and testing, detection performance metrics, and experimental procedure are discussed in this section.

3.1 Logistic Model Tree (LMT) Algorithm

The LMT algorithm is a hybrid of linear logistic regression and the decision tree algorithm. It can generate a model with high predictive precision while still generating an interpretable model. In this research, LMT is used to identify phishing websites, which is a difficult task in cybersecurity. LMT is a hierarchical architecture comprised of a single root, branches, leaves, and nodes. It constructs a standard C4.5 DT with an LR at the node level path down to the leaves. When making a splitting decision, it considers the information gain ratio [25, 26]. These distinguishing characteristics of LMT account for its inclusion as a base learner in this study. Table 1 shows the LMT parameter settings used in this analysis.

Table 1. Classification algorithm

Classification algorithm	Parameter setting
Logistic Model Tree (LMT)	splitOnResiduals = false; useAIC = false; batchSize = 100; fastRegression = True; weightTrimBeta = 0; numBoostingIterations = −1

3.2 Rotation Forest-Based Logistic Model Tree (RF-LMT) Method

Rotation Forest-based Logistic Model Tree (RF-LMT) is a meta-learner that produces classifier models using feature extraction. RF-LMT creates training data for a baseline learner (in this case, LMT) by randomly splitting the feature set into N subsets, and principal component analysis (PCA) is deployed on each of the generated subsets. To maintain the variability in the data, all principal components are kept. Hence, N axis rotations occur to create new features for the baseline learner LMT. The essence of the rotation is to allow concurrent independent accuracy and diversity within the ensemble. Diversity is attained via feature extraction for each baseline learner.

RF-LMT algorithm is presented in Algorithm 1 (See Fig. 1) with the assumption that X is the training dataset, Y is the class label, and F is the feature sets.

3.3 Website Phishing Datasets

Three phishing datasets were used in this study's experimentation phase. These datasets are commonly accessible and are often used in existing studies [1, 11–13, 15]. There

Algorithm 1. RF-LMT Algorithm

Input:
Training set $X = \{x_i, y_i\}, i = 1 \ldots m, y_i \in Y, Y = \{c_1, c_2, \ldots, c_k\}$, c_k is the class label;
Base-Line Learner: *LMT*
$T = 100$ //Iteration count

1. Choose a value for K which is a factor of n, let F randomly divided into K parts of the distinct subsets while each subset must contain $N = n/k$ number of features.
2. Select the corresponding columns of attributes in the subset $T_{i,j}$ from the training dataset X, then form a new matrix $X_{i,j}$. Extract a bootstrap subset of objects ¾ of X to make a new training dataset $X'_{i,j}$.
3. Use Matrix $X'_{i,j}$ as feature transform to produce the co-efficient in the matrix $P_{i,j}$, which j^{th} column coefficient is the characteristic component j^{th}.
4. Construct a sparse rotation matrix S_i using the obtained coefficient obtained in the matrix $P_{i,j}$.
5. Classifier T_i of $d_{i,j}(XS_i^f)$ to determine x belonging to the class y_i, Then, calculate class confidence: $\alpha_j(x) = \frac{1}{L}\sum_{i=1}^{L} d_{i,j}(XS_i^f)$.

Output Assign the category with the largest $\alpha_j(x)$ value to x.

Fig. 1. Pseudocode for proposed RF-LMT method

are 11,055 instances in the first dataset (Dataset A; 4,898 phishing and 6,157 legitimate instances). Dataset A contains 30 distinct attributes that define the dataset [1]. The second dataset (Dataset B) contains 10,000 instances, 5,000 of which are legitimate and 5,000 of which are phishing. Dataset B comprises 48 discrete, continuous, and categorical functions. [11, 13]. The third dataset (Dataset C) comprises 1,353 instances with a total of ten attributes (702 phishing, 548 real, and 103 suspicious). Dataset C is distinguished from Datasets A and B, having three class labels. For more information on the phishing datasets, see [1, 11–13, 15].

3.4 Experimental Procedure

This section presents the experimental procedure as seen in Fig. 2 that was used in this study. The procedure is intended to empirically evaluate and validate the efficacy of the proposed methods for detecting phishing websites. Three phishing datasets from the UCI repositories are used for training and testing the proposed methods. The proposed website phishing detection model is developed and evaluated using K-fold ($k = 10$) Cross-Validation (CV) method. The 10-fold CV selection is based on its ability to create phishing models while minimizing the impact of the class imbalance problem [27, 28]. Since the K-fold CV technique allows each instance to be used iteratively for both training and testing [28–31], the proposed model (RF-LMT) and selected baseline classifiers (Multilayer Perceptron (MLP), K Nearest Neighbour (KNN), Decision Tree (DT), Bayesian Network (BN)) were deployed on phishing datasets based on 10-fold CV. The selected baseline classifiers were chosen based on their usage and performance from existing studies [32–35]. The phishing detection efficiency of the proposed phishing

model (RF-LMT) was then evaluated and compared to other experimented and existing phishing detection approaches. All experiments were performed using the WEKA machine learning tool in the same environment [36].

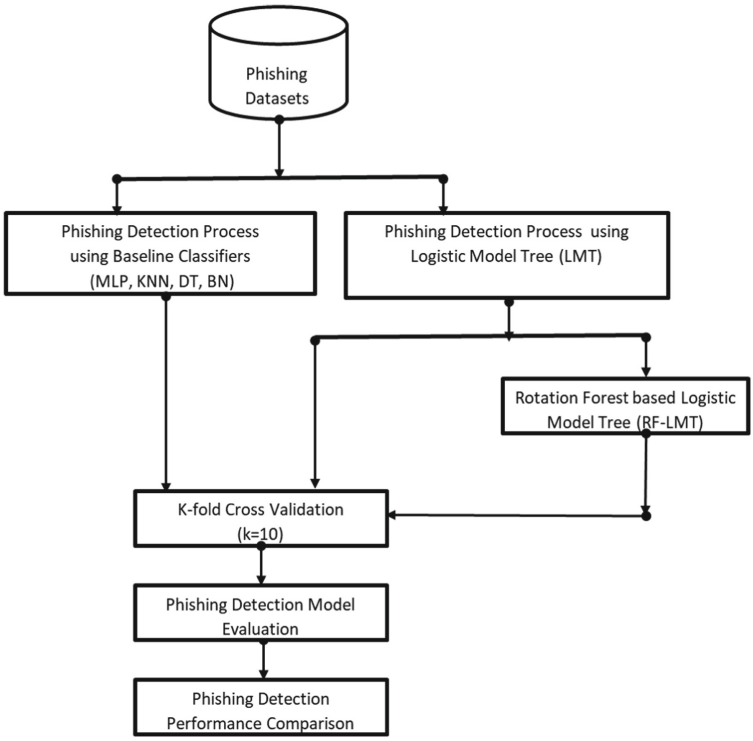

Fig. 2. Experimental procedure

3.5 Performance Evaluation Metrics

Accuracy, F-measure, Area under the Curve (AUC), False-Positive Rate (FPR), True Positive Rate (TPR), and Mathew's Correlation Coefficient (MCC) performance evaluation metrics are used to assess the detection performance of the experimented phishing models. The preference for these metrics stems from the widespread and regular use of these metrics for website phishing detection in existing studies [11, 12, 17–19, 37, 38].

i. Accuracy is the average degree at which the actual labels of all instances are predicted correctly. It is computed as outlined in Eq. (1):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{1}$$

ii. F-measure shows the weighted average of the Recall (R) and Precision (P). It stresses a classifier's ability to maximize both precision and recall at the same time. Equation 2 represents the computation of the F-measure.

$$\text{F} - \text{measure} = \frac{2 \times \text{P}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{2}$$

iii. The AUC plots the FP rate on the X-axis and the TP rate on the Y-axis. AUC is not vulnerable to plurality bias and does not overlook the minority samples during its assessment.

iv. The False Positive Rate (FPR) is the proportion of legitimate instances mistakenly reported as phishing attacks.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \times 100 \tag{3}$$

v. True Positive Rate (TPR) is the rate at which actual phishing website instances are correctly classified as that phishing website.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{4}$$

vi. The Mathews Correlation Coefficient (MCC) is a statistical rate that provides a high score if the prediction produces good outcomes in all four classes of the confusion matrix (true positives, false negatives, true negatives, and false positives), in proportion to the scale of the positive and negative elements in the dataset. MCC can be computed as shown in Eq. 5.

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \tag{5}$$

4 Results and Discussion

This section discusses the experimental findings obtained when the experimental framework was implemented, trained, and tested with three phishing datasets.

4.1 LMT and Baseline Classifiers

As documented in Table 2, the performance of LMT was compared with selected experimented baseline classifiers on Dataset A. Six performance evaluation metrics were used for the performance comparison (See Sect. 3.5). Based on accuracy values, LMT yielded the highest accuracy value of 96.92% when compared with KNN (96.84%), DT (95.87%), MLP (94.76%), and BN (92.98%). Similar performance can be observed in terms of f-measure and AUC values. In particular, LMT recorded a f-measure and AUC values of 0.969 and 0.99 respectively which outperformed KNN (0.968, 0.967),

DT(0.959, 0.984), MLP(0.948, 0.983) and BN(0.93, 0.981). Also, LMT on Dataset A had the highest TP-Rate (0.969) and lowest FP-Rate (0.033) values compared with the baseline classifiers. Although it can be observed that the performance of LMT on Dataset A is comparable to baseline classifiers such as KNN, however, the hyper-parameterization of KNN is a drawback [39].

Table 2. Experimental results of LMT and baseline classifiers on Dataset A

	LMT	MLP	KNN	DT	BN
Accuracy (%)	96.92	94.76	96.84	95.87	92.98
F-Measure	0.969	0.948	0.968	0.959	0.93
AUC	0.990	0.983	0.967	0.984	0.981
TP-Rate	0.969	0.948	0.968	0.959	0.930
FP-Rate	0.033	0.053	0.034	0.045	0.075
MCC	0.938	0.894	0.936	0.916	0.858

Table 3. Experimental results of LMT and baseline classifiers on Dataset B

	LMT	MLP	KNN	DT	BN
Accuracy (%)	97.91	95.92	95.53	97.31	95.79
F-Measure	0.979	0.959	0.955	0.973	0.958
AUC	0.993	0.983	0.955	0.976	0.992
TP-Rate	0.979	0.959	0.955	0.973	0.958
FP-Rate	0.021	0.041	0.045	0.027	0.042
MCC	0.958	0.918	0.911	0.946	0.916

Table 4. Experimental results of LMT and baseline classifiers on Dataset C

	LMT	MLP	KNN	DT	BN
Accuracy (%)	89.36	84.77	86.32	87.58	84.33
F-Measure	0.894	0.840	0.863	0.891	0.828
AUC	0.972	0.927	0.880	0.916	0.948
TP-Rate	0.894	0.848	0.863	0.890	0.843
FP-Rate	0.079	0.108	0.104	0.082	0.118
MCC	0.813	0.742	0.761	0.803	0.727

Correspondingly, on Dataset B, the performance of LMT was superior to the baseline classifiers. As presented in Table 3, LMT achieved the highest accuracy value (97.91%),

F-Measure value (0.979), AUC value (0.993), TP-Rate value (0.979), MCC value (0.958), and the lowest FP-Rate value (0.021) when compared with the performance of the baseline classifiers. Furthermore, similar findings were observed on the performance of LMT on Dataset C, as presented in Table 4. LMT, in most cases, was significantly superior to most of the experimented baseline classifiers. These observations indicate that LMT provided equivalent results (performance) for phishing detection across all three datasets, regardless of dataset size. In other words, LMT showed competitive performance against baseline classifiers in website phishing detection. However, the performance of LMT can be amplified by augmenting it with an appropriate meta-leaner (Rotation Forest) as proposed in this study.

4.2 Rotation Forest-Based Logistic Model Tree (RF-LMT)

In this section, the performance of the proposed RF-LMT with the LMT classifier is presented and compared. Recall from the previous section (See Sect. 4.1), the superiority of the performance of LMT over selected baseline classifiers in website phishing detection has been emphasized. In this context, however, the objective is to see how well the proposed RF-LMT method will perform compared to the LMT classifier. The results of LMT and RF-LMT are presented in Table 5.

Observations from these results indicate that the proposed RF-LMT had promising results and, based on most performance metrics, outperformed the LMT classifier on Dataset A. For instance, RF-LMT recorded an accuracy value of 97.33% as against 96.92% produced by LMT. Also, a similar pattern of improvement can be observed on the evaluation metric, as shown in Table 5. Specifically, RF-LMT had a superior f-measure value (0.973), AUC value (0.997), TP-Rate value (0.973), and MCC value as compared with LMT.

Table 5. Experimental results of RF-LMT and LMT on Dataset A

	LMT	RF-LMT
Accuracy (%)	96.92	97.33
F-Measure	0.969	0.973
AUC	0.990	0.997
TP	0.969	0.973
FP	0.033	0.029
MCC	0.938	0.946

Furthermore, RF-LMT outperformed the LMT classifier on Dataset B and Dataset C based on performance evaluation metrics as used in this study. On Dataset B, RF-LMT achieved an accuracy of 98.24%, F-Measure of 0.982, AUC of 0.998, TP-Rate of 0.982, FP-Rate of 0.018, and MCC of 0.965, respectively, as shown in Table 6. This is better when compared with LMT results which had lower performance. Also,

Table 6. Experimental results of RF-LMT and LMT on Dataset B

	LMT	RF-LMT
Accuracy (%)	97.91	98.24
F-Measure	0.979	0.982
AUC	0.993	0.998
TP	0.979	0.982
FP	0.021	0.018
MCC	0.958	0.965

Table 7. Experimental results of RF-LMT and LMT on Dataset C

	LMT	RF-LMT
Accuracy (%)	89.36	90.61
F-Measure	0.894	0.906
AUC	0.972	0.977
TP	0.894	0.906
FP	0.079	0.068
MCC	0.813	0.835

on Dataset C, a similar pattern of results was observed (See Table 7) as the proposed RF-LMT outperformed the LMT classifier.

Consequently, the superior detection capabilities of RF-LMT on the experimented datasets imply that it has a lower likelihood of misclassifying phishing attacks than LMT. Additionally, the high AUC and MCC values of RF-LMT demonstrate its resistance and resilience to inherent data quality problems such as class imbalance and high dimensionality on the analyzed datasets than LMT. Although LMT performed comparably well and competitive with baseline classifiers such as KNN, MLP, BN, and DT. However, the proposed RF-LMT is better than LMT as the meta-learner (Rotation Forest) improved the performance of LMT. These results are consistent with observations on the application of ensemble techniques in other perspectives [27, 40, 41].

4.3 Rotation Forest-Based Logistic Model Tree (RF-LMT) with Existing Methods

In this section, the performance of the proposed RF-LMT is further compared with existing state-of-the-art methods for website phishing detection. Table 8 shows the performance comparison of RF-LMT with existing methods on Dataset A. Specifically, the experimental results from Al-Ahmadi and Lasloum [42], Alsariera, Elijah and Balogun [12], Ali and Malebary [21], and Vrbančič, Fister Jr and Podgorelec [2] are comparable to that of RF-LMT. However, RF-LMT still outperformed these models in accuracy and other metric values based on Dataset A.

Table 8. Performance evaluation of RF-LMT and existing models on Dataset A

Phishing models	Accuracy (%)	F-Measure	AUC	TP-Rate	FP-Rate	MCC
Aydin and Baykal [14]	95.39	0.938	0.936	–	0.046	–
Dedakia and Mistry [8]	94.29	–	–	–	–	–
Ubing, Jasmi, Abdullah, Jhanjhi and Supramaniam [17]	95.40	0.947	–	–	0.041	–
Hadi, Aburub and Alhawari [10]	92.40	–	–	–	–	–
Chiew, Tan, Wong, Yong and Tiong [13]	93.22	–	–	–	–	–
Rahman, Rafiq, Toma, Hossain and Biplob [11] (KNN)	94.00	–	–	–	0.049	–
Rahman, Rafiq, Toma, Hossain and Biplob [11] (SVM)	95.00	–	–	–	0.039	–
Chandra and Jana [9]	92.72	–	–	–	–	–
Folorunso, Ayo, Abdullah and Ogunyinka [19] (Stacking)	95.97	–	–	–	–	–
Folorunso, Ayo, Abdullah and Ogunyinka [19] (Hybrid NBTree)	94.10	–	–	–	–	–
Al-Ahmadi and Lasloum [42]	96.65	0.965	–	–	–	–
Alsariera, Elijah and Balogun [12]	96.26	–	–	–	0.040	–
Ali and Malebary [21]	96.43	–	–	–	–	–
Ferreira, Martiniano, Napolitano, Romero, Gatto, Farias and Sassi [43]	87.61	–	–	–	–	–
Vrbančič, Fister Jr and Podgorelec [2]	96.50	–	–	–	–	–
***Proposed RF-LMT**	**97.33**	**0.973**	**0.997**	**0.973**	**0.029**	**0.946**

Likewise, Table 9 compared the performance of the proposed method with existing methods based on Dataset B. In particular, the performance of RF-LMT was superior to methods proposed by Chiew, Tan, Wong, Yong and Tiong [13] and Rahman, Rafiq, Toma, Hossain and Biplob [11]**.** Also, based on Dataset C, as shown in Table 10, RF-LMT outperformed existing methods as proposed by Rahman, Rafiq, Toma, Hossain

Table 9. Performance evaluation of RF-LMT and existing models on Dataset B

Phishing Models	Accuracy (%)	F-Measure	AUC	TP-Rate	FP-Rate	MCC
Chiew, Tan, Wong, Yong and Tiong [13]	94.60	–	–	–	–	–
Rahman, Rafiq, Toma, Hossain and Biplob [11] (KNN)	87.00	–	–	–	0.078	–
Rahman, Rafiq, Toma, Hossain and Biplob [11] (SVM)	91.00	–	–	–	0.067	–
Proposed RF-LMT	98.24	0.982	0.998	0.982	0.018	0.965

and Biplob [11]. These findings further show the superiority of the proposed RF-LMT as it in most cases outperformed existing website phishing methods based on multiple phishing datasets.

Table 10. Performance evaluation of RF-LMT and existing models on Dataset C

Phishing Models	Accuracy (%)	F-Measure	AUC	TP-Rate	FP-Rate	MCC
Rahman, Rafiq, Toma, Hossain and Biplob [11] (KNN)	88.00	–	–	–	0.099	–
Rahman, Rafiq, Toma, Hossain and Biplob [11] (SVM)	87.00	–	–	–	0.087	–
Proposed RF-LMT	90.61	0.906	0.977	0.906	0.068	0.835

Conclusively, the Research Questions (RQs) posed in the introduction were examined at the end of the experimentation. The following conclusions were reached:

RQ1: *How efficient is the LMT in detecting legitimate and phishing websites?*

LMT algorithm implementations indeed produced significant improvement as compared with baseline methods such as MLP, KNN, DT, and BN with better accuracy and other performance evaluation metrics. This performance is replicated across the three datasets that were considered in this study.

RQ2: *How efficient is the proposed RT-LMT algorithm in detecting legitimate and phishing websites?*

As compared to LMT for phishing website detection, the proposed RT-LMT leveraged the promising success of LMT and demonstrated a substantial increase in accuracy as well as a decrease in error rate. This progress was repeated and observed across the experimented three datasets.

RQ3: *How efficient is the proposed RF-LMT compared to existing phishing methods?*

The performance of the proposed RF-LMT is superior in terms of accuracy, F-Measure, AUC, TP-Rate, FP-Rate, and MCC values as used in this study compared with existing state-of-the-art methods using the three datasets for phishing website detection.

5 Conclusion and Future Works

Phishing attacks are one of the severe cyberattacks that have a global negative effect on internet users. A website phishing attack can be harmful to internet users and internet-based solutions in general. A website phishing attack helps an adversary access victims' personal information, which can then be used to conduct fraudulent transactions or capture users' identities. However, due to attackers' advanced and dynamic strategies, identifying phishing websites has proven difficult. Hence, this study proposed RF-LMT that leveraged the performance of the LMT classifier to detect phishing websites. RF-LMT recorded superior detection performance that outperformed **baseline** models such as MLP, KNN, DT, BN, and existing state-of-the-art methods for phishing website detection.

The authors plan to test the proposed RF-LMT on additional real-time phishing website datasets in the future to determine its generalization potential in detecting phishing websites. Also, more sophisticated models for developing scalable models will be investigated.

References

1. Mohammad, R.M., Thabtah, F., McCluskey, L.: Predicting phishing websites based on self-structuring neural network. Neural Comput. Appl. **25**(2), 443–458 (2013). https://doi.org/10.1007/s00521-013-1490-z
2. Vrbančič, G., Fister Jr, I., Podgorelec, V.: 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 (2018)
3. Ali, W., Ahmed, A.A.: Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting. IET Inf. Secur. **13**, 659–669 (2019)
4. Verma, R., Das, A.: 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, pp. 55–63 (2017)
5. Azeez, N., Misra, S., Margaret, I.A., Fernandez-Sanz, L.: Adopting automated whitelist approach for detecting phishing attacks. Comput. Secur. **108**, 102328 (2021)
6. Alqahtani, M.: Phishing Websites Classification using Association Classification (PWCAC). In: 2019 International Conference On Computer and Information Sciences (ICCIS), pp. 1–6. IEEE (2019)
7. Abdelhamid, N., Ayesh, A., Thabtah, F.: Phishing detection based associative classification data mining. Expert Syst. Appl. **41**, 5948–5959 (2014)
8. Dedakia, M., Mistry, K.: Phishing detection using content based associative classification data mining. J. Eng. Comput. Appl. Sci. **4**, 209–214 (2015)
9. Chandra, Y., Jana, A.: Improvement in phishing websites detection using meta classifiers. In: 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 637–641. IEEE (2019)
10. Hadi, W.e., Aburub, F., Alhawari, S.: A new fast associative classification algorithm for detecting phishing websites. Appl. Soft Comput. **48**, 729–734 (2016)

11. Rahman, S.S.M.M., Rafiq, F.B., Toma, T.R., Hossain, S.S., Biplob, K.B.B.: Performance assessment of multiple machine learning classifiers for detecting the phishing URLs. In: Raju, KSrujan, Senkerik, R., Lanka, S.P., Rajagopal, V. (eds.) Data Engineering and Communication Technology. AISC, vol. 1079, pp. 285–296. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-1097-7_25
12. Alsariera, Y.A., Elijah, A.V., Balogun, A.O.: Phishing website detection: forest by penalizing attributes algorithm and its enhanced variations. Arab. J. Sci. Eng. **45**(12), 10459–10470 (2020). https://doi.org/10.1007/s13369-020-04802-1
13. Chiew, K.L., Tan, C.L., Wong, K., Yong, K.S., Tiong, W.K.: A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. Inf. Sci. **484**, 153–166 (2019)
14. Aydin, M., Baykal, N.: Feature extraction and classification phishing websites based on URL. In: 2015 IEEE Conference on Communications and Network Security (CNS), pp. 769–770. IEEE (2015)
15. Adeyemo, V.E., Balogun, A.O., Mojeed, H.A., Akande, N.O., Adewole, K.S.: Ensemble-based logistic model trees for website phishing detection. In: Anbar, M., Abdullah, N., Manickam, S. (eds.) ACeS 2020. CCIS, vol. 1347, pp. 627–641. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-6835-4_41
16. Pham, B.T., Nguyen, V.-T., Ngo, V.-L., Trinh, P.T., Ngo, H.T.T., Bui, D.T.: A novel hybrid model of rotation forest based functional trees for landslide susceptibility mapping: a case study at Kon Tum Province, Vietnam. In: Bui, D.T., Do, A.N., Bui, H.-B., Hoang, N.-D. (eds.) GTER 2017, pp. 186–201. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-68240-2_12
17. Ubing, A.A., Jasmi, S.K.B., Abdullah, A., Jhanjhi, N., Supramaniam, M.: Phishing website detection: an improved accuracy through feature selection and ensemble learning. Int. J. Adv. Comput. Sci. Appl. **10**, 252–257 (2019)
18. Abdulrahaman, M.D., Alhassan, J.K., Adebayo, O.S., Ojeniyi, J.A., Olalere, M.: (2019): Phishing attack detection based on random forest with wrapper feature selection method. Int. J. Inf. Process. Commun. **7**, 209–224 (2019)
19. Folorunso, S.O., Ayo, F.E., Abdullah, K.-K.A., Ogunyinka, P.I.: Hybrid vs ensemble classification models for phishing websites. Iraqi J. Sci. 3387–3396 (2020). https://doi.org/10.24996/ijs.2020.61.12.27
20. Alsariera, Y.A., Adeyemo, V.E., Balogun, A.O., Alazzawi, A.K.: Ai meta-learners and extra-trees algorithm for the detection of phishing websites. IEEE Access **8**, 142532–142542 (2020)
21. Ali, W., Malebary, S.: Particle swarm optimization-based feature weighting for improving intelligent phishing website detection. IEEE Access **8**, 116766–116780 (2020)
22. Osho, O., Oluyomi, A., Misra, S., Ahuja, R., Damasevicius, R., Maskeliunas, R.: Comparative evaluation of techniques for detection of phishing URLs. In: Florez, H., Leon, M., Diaz-, J.M., Belli, S. (eds.) Applied Informatics: Second International Conference, ICAI 2019, Madrid, Spain, November 7–9, 2019, Proceedings, pp. 385–394. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-32475-9_28
23. Balogun, A.O., Basri, S., Abdulkadir, S.J., Adeyemo, V.E., Imam, A.A., Bajeh, A.O.: Software defect prediction: analysis of class imbalance and performance stability. J. Eng. Sci. Technol. **14**, 3294–3308 (2019)
24. Yu, Q., Jiang, S., Zhang, Y.: The performance stability of defect prediction models with class imbalance: an empirical study. IEICE Trans. Info. Sys. **100**, 265–272 (2017)
25. Lee, S., Jun, C.-H.: Fast incremental learning of logistic model tree using least angle regression. Expert Syst. Appl. **97**, 137–145 (2018)
26. Sumner, M., Frank, E., Hall, M.: Speeding up logistic model tree induction. In: Jorge, A.M., Torgo, L., Brazdil, P., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 675–683. Springer, Heidelberg (2005). https://doi.org/10.1007/11564126_72

27. Balogun, A.O., et al.: SMOTE-based homogeneous ensemble methods for software defect prediction. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12254, pp. 615–631. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58817-5_45
28. Yadav, S., Shukla, S.: Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 78–83. IEEE (2016)
29. Arlot, S., Lerasle, M.: Choice of V for V-fold cross-validation in least-squares density estimation. J. Mach. Learn. Res. **17**, 7256–7305 (2016)
30. Balogun, A.O., et al.: Search-based wrapper feature selection methods in software defect prediction: an empirical analysis. In: Silhavy, R. (ed.) CSOC 2020. AISC, vol. 1224, pp. 492–503. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51965-0_43
31. Basri, S., Almomani, M.A., Imam, A.A., Thangiah, M., Gilal, A.R., Balogun, A.O.: The organisational factors of software process improvement in small software industry: comparative study. In: Saeed, F., Mohammed, F., Gazem, N. (eds.) IRICT 2019. AISC, vol. 1073, pp. 1132–1143. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-33582-3_106
32. Ahmad, S.N.W., Ismail, M.A., Sutoyo, E., Kasim, S., Mohamad, M.S.: Comparative performance of machine learning methods for classification on phishing attack detection. Int. J. **9**, 349–354 (2020)
33. Jain, A.K., Gupta, B.: Comparative analysis of features based machine learning approaches for phishing detection. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 2125–2130. IEEE (2016)
34. Karabatak, M., Mustafa, T.: Performance comparison of classifiers on reduced phishing website dataset. In: 2018 6th International Symposium on Digital Forensic and Security (ISDFS), pp. 1–5. IEEE (2018)
35. Balogun, A.O., et al.: Empirical analysis of rank aggregation-based multi-filter feature selection methods in software defect prediction. Electronics **10**, 179 (2021)
36. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. ACM Sig. Exp. **11**, 10–18 (2009)
37. Adewole, K.S., Akintola, A.G., Salihu, S.A., Faruk, N., Jimoh, R.G.: Hybrid rule-based model for phishing URLs detection. In: Miraz, M.H., Excell, P.S., Ware, A., Soomro, S., Ali, M. (eds.) Emerging Technologies in Computing: Second International Conference, iCETiC 2019, London, UK, August 19–20, 2019, Proceedings, pp. 119–135. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-23943-5_9
38. AlEroud, A., Karabatis, G.: Bypassing Detection of URL-based phishing attacks using generative adversarial deep neural networks. In: Proceedings of the Sixth International Workshop on Security and Privacy Analytics, pp. 53–60 (2020)
39. Mabayoje, M.A., Balogun, A.O., Jibril, H.A., Atoyebi, J.O., Mojeed, H.A., Adeyemo, V.E.: Parameter tuning in KNN for software defect prediction: an empirical analysis. Jurnal Teknologi dan Sistem Komputer **7**, 121–126 (2019)
40. Adeyemo, V.E., Azween, A., JhanJhi, N., Mahadevan, S., Balogun, A.O.: Ensemble and deep-learning methods for two-class and multi-attack anomaly intrusion detection: an empirical study. Int. J. Adv. Comput. Sci. Appl. **10**, 520–528 (2019)
41. Balogun, A.O., Balogun, A.M., Sadiku, P.O., Amusa, L.: An ensemble approach based on decision tree and Bayesian network for intrusion detection. Ann. Comput. Sci. Ser. **15**, 82–91 (2017)
42. Al-Ahmadi, S., Lasloum, T.: PDMLP: phishing detection using multilayer perceptron. Int. J. Netw. Secur. Appl. **12**, 59–72 (2020)
43. Ferreira, R.P., et al.: Artificial neural network for websites classification with phishing characteristics. Soc. Netw. **7**, 97 (2018)