Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Summer and Autumn 2021, 809-823
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2021.5457

Recognizing phishing websites based on a bayesian combiner

Omid Rahmani Seryasat [a,*], Sina Ahmadi[b], Pouya Yousefi[b], Farzad Tat Shahdost[c], Sareh Sanei[d]

[a]*Assistant Professor, Department of Electrical Engineering, Shams Higher Education Institute, Iran.*
[b]*Department of Computer Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran.*
[c]*Department of Electrical Engineering, Islamic Azad University, Garmsar Branch, Semnan, Iran.*
[d]*Department of Electrical Engineering, Technical and Vocational University (TVU), Tehran, Iran.*

(Communicated by Ehsan Kozegar)

Abstract

Phishing is a social engineering technique used to deceive users, which means trying to obtain confidential information such as username, password or bank account information. One of the most important challenges on the Internet today is the risk of phishing attack and Internet scams. These attacks cost the United States billions of dollars a year. Therefore, researchers have made great efforts to identify and combat such attacks. Accordingly, the present study aims to evaluate the methods of identifying phishing websites. This research is applied in terms of its objectives and descriptive-analytical in nature. In this article, the classification approach is used to identify phishing websites. From a machine learning point of view, if a suitable strategy is used, the ensemble of votes of different classifiers can be used to increase the accuracy of classification. In the method proposed in this paper, three inherently different ensemble classifiers, called bagging, AdaBoost, and rotation forest are employed. In this method, the stacked generalization strategy is used as an ensemble strategy. A relatively new dataset is employed to evaluate the performance of the proposed method. The database was added to the UCI Database in 2015 and uses 30 features that appear to be appropriate for distinguishing phishing and non-phishing websites. The present study uses 10-fold-cross-validation method as an evaluation strategy. The numerical results indicate that the proposed method can be used as a promising method for detecting phishing websites. It is worth mentioning that in this method, an F-score of 96.3 is resulted, which is a good result in detecting phishing.

Keywords: Phishing, Classification, Ensembling, Stacked generalization

*Corresponding author
Email addresses: orseryasat@shamsgonbad.ac.ir (Omid Rahmani Seryasat), s.ahmadi.engr@gmail.com (Sina Ahmadi), yousefi.pouya@wtiau.ac.ir (Pouya Yousefi), Farzad_tat@yahoo.com (Farzad Tat Shahdost), sareh-sanei@tvu.ac.ir (Sareh Sanei)

Received: July 2021 *Accepted:* September 2021

1. Introduction

Machine learning methods have a wide range of applications [9, 17]. Phishing is a type of computer attack in which an attacker uses electronic communication channels to communicate with humans and uses social engineering messages to persuade them to do things that benefit the attacker [8].

One of the most important challenges in Internet banking today is the risk of phishing attacks and Internet scams, which causes a lot of damage to customers and organizations. In addition to lost money and time, individuals' trust in online programs and services is lost, and as a result, the credibility of organizations is diminished. According to Gartner [3], these attacks cause several billion dollars damage annually in the United States alone [2].

Therefore, experts and researchers have made great efforts to identify and combat such attacks. Many tools have been developed to identify and deal with them, but since Fishers are constantly changing their way of working at low cost, these tools need to be updated to identify their methods.

According to Symantec [21], the rate of phishing attacks has increased in 2013, so that in 2012 this rate was one phishing email out of every 414 emails, while in 2013 it was one phishing email for every 392 emails. Has been reported. Also, the rate of phishing attacks in May 2014 was reported as one phishing email for every 395 emails. The growth rate of these attacks is such that the need to identify and deal with them is felt more than ever. Today, despite the design of many tools to detect and combat these types of attacks, as well as the spread of public awareness about phishing, these attacks are still a serious threat on the Internet and their number is increasing day by day.

Phishing detection methods are divided into two categories: user awareness and software detection [8]. The focus of this study is on software methods. These methods are divided into 4 general categories: 1) blacklist-based methods, 2) metaheuristic methods, 3) appearance similarity, and 4) machine learning. In this article, a method that falls into the fourth category has been employed. Phishing website identification techniques in the machine learning category look for solutions for document clustering or classifying. In these methods, models are made using machine learning and clustering algorithms. These algorithms include the nearest neighbor, C4.5, the DBSCAN support vector machine, and k-means. Among the methods presented in this category are: textual and visual anti-phishing with Bayesian approach [20], large-scale automatic classification of pages [23], and Bayesian Anti-Phishing Toolbar (B-APT) [11].

Liu et al. [12] point out that it is possible to identify phishing websites as well as their targets by finding websites that look like suspicious pages. If a suspicious website is similar to a website with a different domain name, then the suspicious website is considered a phishing website. For example, if a website is very similar to Paypal, it is definitely a phishing website aimed at attacking Paypal. This method is based on data mining and uses the classification technique. Table 1 summarizes the detection methods of phishing sites.

2. Suggested method

In this section, the ensemble classification methods used in this study are first examined. These include the Bagging, AdaBoost, and Rotation Forest classifiers. After introducing these methods, we will use a new method to combine these strong classifiers to increase the discriminability power of the classifier.

2.1. Bagging

The idea of Bagging (Bootstrap AGGregatING) is simple and clear. In this method, a combination of classifiers is created, each of which is trained on a bootstrap built from the original data set [1]. Then, the final vote will be the vote of the majority of the categories. Suppose $Z = \{z_1, \ldots, z_N\}$ is

Table 1: Summary of phishing attack detection methods

No.	Method	Basis	Checking the A data section	Checking the B data section section	Checking the C data section
1	PhishNet	Blacklist	•	•	
2	SpoofGuard	Metaheuristic		•	•
3	CID		•	•	•
4	PhishGuard			•	•
5	Cantina			•	•
6	Blacklist generator			•	•
7	Phishing page identification and goal detection			•	•
8	URL-based detection			•	•
9	Preventing phishing attacks by extracting page rank, credibility, and source code			•	•
10	Detection based on appearance similarity without having victim website information	Appearance similarity		•	•
11	Combat phishing with key distinct features			•	•
12	Textual and visual anti-phishing: a Bayesian approach	Machine learning		•	•
13	large-scale automatic classification of pages		•	•	•
14	Bayesian Anti-Phishing Toolbar (B-APT)			•	•
15	Automatic detection of phishing goal from the phishing page			•	•

the original data set and we want to build a bootstrap from them. We select n samples of this set randomly and by placement, and thus we make the first training set for the first classifier, then we repeat the same thing for the second classifier, for which a training set is obtained from the Z set. So we will have L classifiers and L training sets. The classifiers are trained on the training sets related to them. Now, when we provide a sample as a test to the ensemble classifier, the voting mechanism is

used to determine the label of this sample, and the vote of the majority of the classifiers will indicate the class of the test sample.

2.2. AdaBoost

In AdaBoost (Adaptive Boosting), a number of weak classifiers are combined to form a strong one [1]. In this article, the constituents of AdaBoost are a number of weak classifiers called Naive Bayes. The purpose of Adaptive Boosting is to increase the weight and selection chance of those samples that are difficult and closer to the classification boundary that the classifiers have difficulty facing them. To better understand this algorithm, see Figure 1. In Figure $1-A$, weak classifier 1 is first applied on the data set and incorrectly classifies the 3 samples. In the next step, the weight of the samples that were misclassified increases and their chances of being selected as elements of the next training set increase (Figure $1-B$). Figure $1-C$ assumes that samples that were incorrectly classified by classifier 1 are provided as training elements of the weak classifier 2. It is noticed that classifier 2 corrects the error of the previous classifier, but it itself has difficulty in classifying other elements. Then the weight of these elements is also increased (Figure $1-D$) and are correctly classified by the weak classifier 3 (Figure $1-E$). Finally, from the combination of these three weak classifiers, a very strong classifier is made that classifies all the samples correctly.

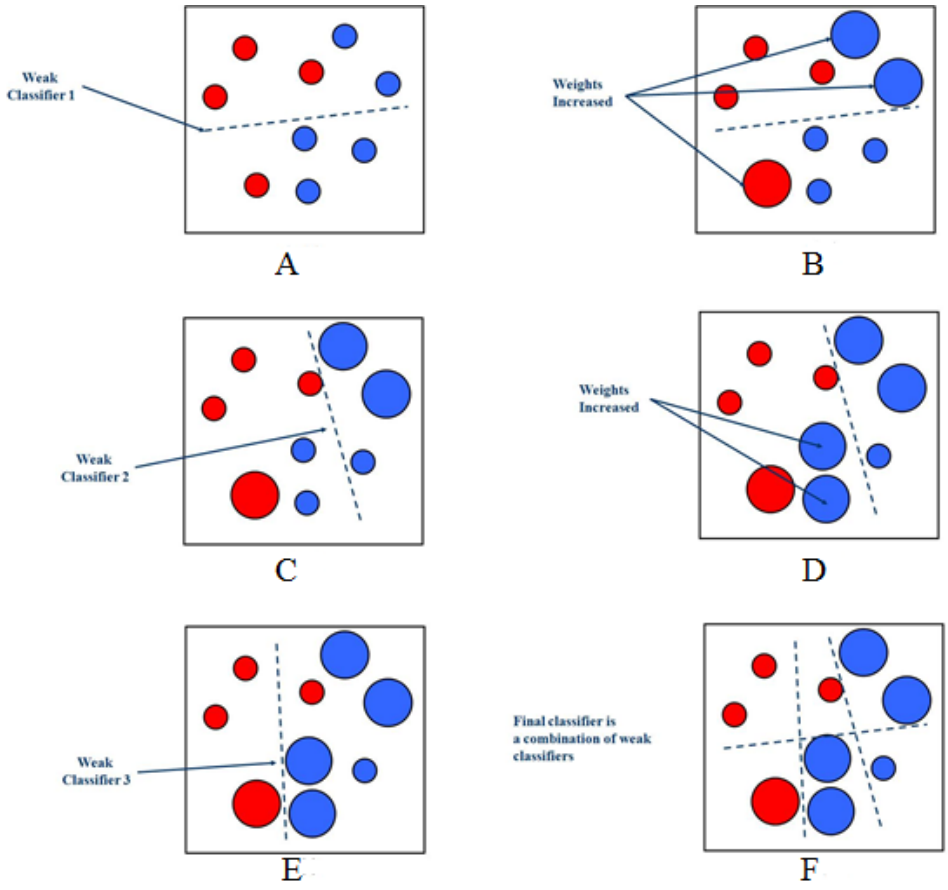

Figure 1: An example of AdaBoost functioning

2.3. Rotation forest

Now let's look at a strong ensemble classifier called rotation forest. Suppose F is the feature set of the classifier problem and the number of classifiers. To build a training set, the following is done:

1. Divide F into k subsets (k is the input of the algorithm). Subsets can be a partition of the main set or shared. To increase diversity, partition sets have been selected in this method. For simplicity, assume that n is a coefficient of k, in which case each subset will have a feature of $M = n/k$.

2. Then, F_{ij} shows the j^{th} subset of features that have been used to teach the di classifier. For each of these subsets, select a non-empty subset of the classes and perform a sampling by placing 75% of the data on it. Run PCA on the selected data and only on M and save the eigenvectors of $a_{i,j}^{1}, \ldots, a_{i,j}^{(M_1)}$ that are each $M \times 1$. Running PCA on part of a class instead of all of them is an attempt to avoid creating the same eigenvectors for different classifiers for which the selected features may be the same.

1. Place the obtained eigenvectors in a solitude matrix (rotation matrix) as follows.

$$\begin{bmatrix} a_{i,1}^{(1)}, a_{i,1}^{(1)}, \ldots, a_{i,1}^{(M_1)} & [0] & \ldots & [0] \\ [0] & a_{i,2}^{(1)}, a_{i,2}^{(2)}, \ldots, a_{i,2}^{(M_2)} & \ldots & [0] \\ \vdots & \vdots & \ddots & \vdots \\ [0] & [0] & \ldots & a_{i,k}^{(1)}, a_{i,k}^{(2)}, \ldots, a_{i,k}^{(M_k)} \end{bmatrix}$$

2. To calculate the training set for classifier d_i, the columns of the rotation matrix (features) must be moved in the order in which they are displayed in the training set, where the obtained matrix is displayed as R_i^{α} and the training set will be XR_i^{α} for the d_i classifier. Below is a pseudocode for the algorithm. Decision trees are sensitive to the rotation of feature axes and can therefore be a good choice as a base classifier. In the following, the training method of this classifier is described:

Having the following elements:
X :training data set (an $N * n$ matrix)
Y : training set class label (an $N * 1$ matrix)
L : number of ensemble classifiers
K : number of subsets
$\{w_1, w_2, \ldots, w_c\}$: A set of class labels,
Prepare the R_i^{α} rotation matrix for $i = 1, \ldots, L$ as follows:
Divide F (feature set) into k subsets: $(j = 1, \ldots, K\ )F_{i,j}$
For $j = 1, \ldots, K$
Suppose $X_{i,j}$ is the X data set for $F_{i,j}$ feature set.
Randomly delete a number of $X_{i,j}$ classes.
Make a "placement random selection" equal to 75% of the available samples in $X_{i,j}$. We call this new set $X_{i,j}^{'}$.
Apply PCA algorithm on $X_{i,j}^{'}$ and obtain its eigenvectors ($C_{i,j}$).
Place $C_{i,j}$ of $j = 1, \ldots, K$ inside the R_i matrix.
Arrange the rotation matrix column in the order of the features in the feature set.
Train the D_i classifier with the (XR_i^{α}, Y) training set.
Now, suppose we want to test a sample like x. Suppose $d_{i,j}(XR_i^{\alpha})$ is the probability that the D_i classifier provided for x belonging to class W_j. In this case, using the following averaging, we select

the class that has the highest value of μ.

$$\mu_j(x) = \frac{1}{L}\sum_{i=1}^{L} d_{i,j}(XR_i^{\alpha}), \quad j = 1, \ldots, K$$

2.4. How to make a trainable combiner

In this section, one of the most important innovations of this article, which is the design of a new ensemble classifier, is discussed. Basically, the design of an ensemble classifier is based on a principle called diversity. This means that somehow we have to diversify the basic classifiers used. This diversity can be done in 4 levels: 1) data level, 2) feature level, 3) classifier level, and 4) combiner level. Diversity at the data level means that the base classifiers should use different samples for training, such as Bagging and AdaBoost. Diversity at the feature level means that not all base classifiers are trained on a fixed set of features and different feature sets should be used for training. The third level is based on the principle that we use different specialists (classifiers) for training. In fact, classifiers with different training methods should be employed. The last level is related to the combiner. For example, several types of strategies can be combined, such as majority voting, weighted voting, and the use of a learner, thus creating diversity. In this article, it is intended to use a new strategy to combine classifiers.

Here, the stack generalization method is employed to train the proposed classifier. The idea of stack generalization is as follows: suppose Z is the name of our dataset that contains N samples labeled 0 and 1. Label 0 indicates that the sample site is safe, and label 1 indicates that the site is a scammer. In the proposed method, the Z dataset is partitioned into 4 separate sets of A, B, C, and D. It is also assumed that all three classifiers introduced in the previous section are used. In the stack generalization method, each of the classifiers is trained using the standard 4-fold cross validation method. At the end of the training procedure according to Figure 2, there are 4 copies of each classifier, each of which is trained on 4 data sets (ABC), (BCD), (ACD) and (ABD).

	A	B	C	D
Training	B	C	D	A
	C	D	A	B
Testing	D	A	B	C

Figure 2: 4-fold cross validation representation

Next, a separate classifier is needed to learn how classifiers vote. To this end, for each sample in subset A, the outputs generated by the classifiers trained on (BCD) subset are retained and used as new features to construct the desired dataset for the combiner. Thus, the 3 outputs generated from the base AdaBoost classifiers (Bagging and rotation forest) and the actual sample label in A form a new feature vector. With these interpretations, the training data set for the combiner classifier is a 4-element binary vector. In the following, the same process is repeated for the samples in subset B, and using the output of the trained classifiers on (ACD), a series of 4-element binary vectors are added to the combiner training set. The same process is repeated for subsets C and D. Once the

combiner training set is complete, we need to train our combiner on this data set. In this article, the simple Bayesian classifier is used for his purpose.

After training the combiner, the 4 subsets are re-integrated into the Z set and the basic classifiers are re-trained. In this way, both the base classifiers and the Bayesian ensemble classifier are prepared to predict the labeling of test samples. As can be seen in Figure 3, when a sample X is prepared for testing, the trained classifiers give their vote on sample X to the combiner. Thus, a 3-element binary vector is assigned to the ensemble classifier to determine the final output.

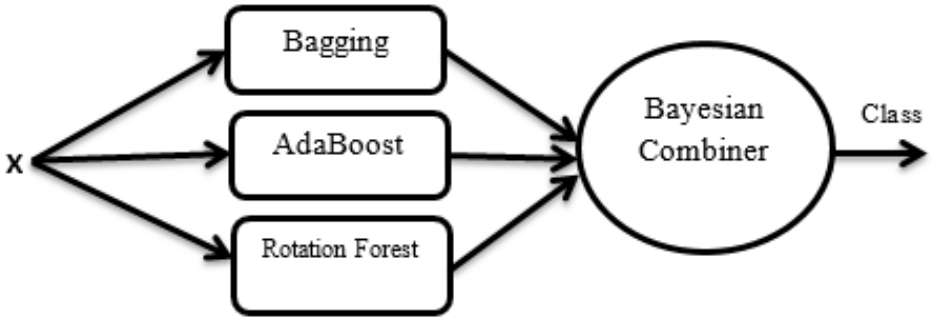

Figure 3: The proposed ensemble classifier including 3 basic classifiers and a trainable combiner

3. Results

This chapter will first describe the database used in this article. Then, the results obtained from the proposed algorithm are reviewed. In addition, the analysis section discusses the effect of using an ensemble classifier instead of individual classifiers.

3.1. Dataset

One of the challenges of research methods working to identify phishing websites is the lack of a standard database. Although many articles are being researched on the prediction of such websites these days, there is no consensus in the literature on the characteristics of fraudulent websites. Mohammad Rami et al. [18] added a database to the UCI data repository that uses 30 features that appear to be suitable for distinguishing phishing and non-phishing websites. In the following, we will categorize these features and briefly explain each of them.

3.1.1. Address bar-based features

- Use of IP address

  When an IP address is used as a substitute for a domain name in a url (such as $http$: $//125.98.3.123/fake.html$), users can be somewhat confident that someone is trying to steal their personal information. Sometimes the IP address is changed to code at base sixteen. For example, $http$: $//0x58.0xCC.0xCA.0x62/2/paypal.ca/index.html$. In general, when the IP address is used instead of the domain name, the site should be suspected of being healthy.

- Use of long URLs to hide suspicious sections

  Scammers can use long URLs to hide suspicious sections in the address bar. For example:

  http://federmacedoadv.com.br/3f/aze/ab51e2e319e51502f416dbe46b773a5e/?cmd = _

  home&dispatch=11004d58f5b74f8dc1e7c2e8dd4105e811004d58f5e5

So URL length can be used as an effective feature to detect fraudulent websites. It should be noted that the average length of URLs in the database used in this article is 54 characters.

- Use of URL shortening services

  URL shortening is a method in which the URL length is significantly reduced, but still leads to the target website. The shortening operation is performed by redirecting to the shortened domain name. In fact, the abbreviated name acts as an interface and links to the long name. For example, the address http://portal.hud.ac.uk/ can be shortened to bit.ly/19DXSk4.

- URLs that have an @ sign

  Using @ in the URL causes the browser to ignore anything before this sign, and the actual address comes after this sign. Therefore, the presence of this sign in the URL can be considered as a suspicious factor.

- Redirect using //

  The presence of // in a URL indicates that the user is being redirected to another website. As an example of such scams, consider the following address:

  http://www.legitimate.com//http://www.phishing.com

  Therefore, the situation where the // sign appears in the URL can indicate whether the destination website is fraudulent or not. In URLs that start with HTTP, this symbol appears in the sixth place, and in those that start with HTTPS, it is in the seventh place. Therefore, if such a rule is not observed in the URLs, the website in question should be suspected.

- Adding a prefix or suffix with the "-" character in the domain

  The hyphen is rarely used in healthy URLs. Scammers often tend to use prefixes or extensions that are separated by a "-" sign in the domain name to give users the feeling that they are using a secure website. For example, look at the following address:

  http://www.Confirme-paypal.com/

- Subdomains and several subdomains

  Suppose we have the following address:

  http://www.hud.ac.uk/students/

  At the address above, uk represents the country code and ac indicates that the website is academic. The combination of the two, ac.uk, is called a level 2 domain, and hud is the real name of the domain. For legal production to extract this feature, first remove www. from the domain name because it is a subdomain. Then, we delete the country code from the address, if any. Finally, we count the number of dots. If the number of dots is more than one, the relevant website is considered suspicious. If the number of these dots is more than 2, then the relevant website can be considered as a fraudulent website because it contains several subdomains.

- HTTPS

  Having HTTPS is very important but not enough to determine if a website is secure. In addition to the HTTPS, the issuer of the HTTPS certificate and the time elapsed since the certificate was issued are also important.

- Domain registration time

  Since phishing websites are created for a short period of time, the number of years it takes for a domain to be registered can indicate a degree of reliability for the relevant website. According to studies on the current dataset, the maximum interval that a site has been used for fraud has been a one-year interval.

- Favicon

  Favicon is a visual icon associated with a website. Many existing user agents, such as graphic browsers and newsreaders, insert Favicon in the address bar as a visual reminder of the website ID. If favicon is loaded from another external domain instead of from the address bar, then the website is likely to be a fraudulent website.

- Use of a non-standard port

  This feature is useful in validating whether a particular service (such as HTTP) is working properly on a particular server or not. In order to control intrusions, it is strongly recommended that only the required ports be included. Therefore, firewalls, proxy servers, and network address translation servers by default leave only a handful of ports open and block the rest. If all ports are open, scammers can take advantage of this and steal users' information. The most important ports and their proper status are shown in Table 2.

Table 2: Common ports to consider

PORT	Service	Meaning	Preferred Status
21	FTP	Transfer files from one host to another	Close
22	SSH	Secure File Transfer Protocol	Close
23	Telnet	provide a bidirectional interactive text-oriented communication	Close
80	HTTP	Hyper test transfer protocol	Open
443	HTTPS	Hypertext transfer protocol secured	Open
445	SMB	Providing shared access to files, printers, serial ports	Close
1433	MSSQL	Store and retrieve data as requested by other software applications	Close
1521	ORACLE	Access oracle database from web.	Close
3306	MySQL	Access MySQL database from web.	Close
3389	Remote Desktop	allow remote access and remote collaboration	Close

- Existence of the word HTTPS in the URL domain

  Scammers often add the word https to the domain of an address to mislead users. For example: http://https-www-paypal-it-webapps-mpp-home.soft-hair.com

3.1.2. Features based on abnormality

- Request URL

  The Request URL checks whether external objects on the page, such as images, videos, and sounds, are loaded from another domain. In secure pages, the page address and most of the objects in it share the same domain.

- URL of Anchor

  An anchor is an element defined by the $< a >$ tag. This element is treated exactly like the request URL. However, for this feature, we consider the following.

  1) If the $< a >$ tags and the website have different domain names, then it will be the same as the request URL.

  2) If the anchor is not connected to any website. As:

  $A. < ahref =, , \#\# >$

  $B. < ahref = "\#content" >$

  $C. < ahref = "\#skip" >$

  $D. < ahref = "JavaScript :: void(0)" >$

- Existence of links in $< meta >, < script >$ and $< link >$ tags

  Secure pages usually have high tags. For example, they use the $< meta >$ tag for metadata in HTML documents. The $< script >$ tag is used to create a script on the client side and the $< link >$ tag is used to retrieve other web resources. These tags are expected to link to the same domain as the website.

- Server Form Handler (SFH)

  SFHs that contain an "about:blank" string are suspicious because action must usually be taken on the information sent. In addition, if the domain name in SFH differs from the web page domain name, then this webpage will be suspicious because the information sent is rarely managed by external domains.

- Send information to email

  Forms on the website allow users to enter their information. This information is usually processed by the servers on which the site is located. Sometimes, fraudsters redirect users' personal information to their personal emails. For this purpose, a server-side scripting language (such as the mail() function in PHP) may be used. Another function on the customer side (such as the mailto: function) may be used. Therefore, using such functions to email customer information can be a reason for websites to be suspicious.

- Unusual URL

  This feature can be extracted from the WHOIS database. For a secure website, the ID is usually part of the website. Therefore, if the host names are not in the URL, the website can be suspected.

3.1.3. HTML and javascript based features

- Forward the website

  One of the most effective features for identifying phishing websites is the number of redirects to the website. Websites that have been redirected more than once are usually suspicious.

- Status bar customization

  Scammers typically use JavaScript to show users a fake URL in the status bar. To extract this feature, we must have access to the source code information of the website. In particular, we need to check whether the onMouseOver event changes the status bar.

- Disabled right-click

  Scammers use functions to disable right-click to prevent users from viewing the page source code. This feature works just like onMouseOver to hide links. Therefore, in the database used in this article, the phrase event.button==2 was searched in the source code and checked whether the right click is disabled or not.

- Use of the pop-up window

  One of the tricks that scammers use to steal users' information is to use a pop-up window to send personal information. Secure websites, on the other hand, use pop-ups for greetings and things like that, not for the user to fill out information.

- IFrame redirection

  IFrame is an HTML tag used to display an additional page within the page being displayed. Scammers can use the IFrame tag and hide it (i.e. without frame borders). Therefore, the use of IFrame can be questionable and can be used as a feature along with other features to detect fraudulent websites.

3.1.4. Domain-based features

- Domain age

  This feature can be extracted from the WHOIS database. Most phishing websites will be available for a short period of time. According to studies, healthy websites are at least 6 months old.

- DNS record An empty DNS record can be a reason for a website to be suspicious because healthy websites usually have this record.

- Website traffic

  This feature evaluates the popularity of a website in terms of the number of visitors and the number of pages they visit. Therefore, since fraudulent websites are available for a short period of time, they are not detected through the Alexa database. According to studies, usually in the worst case, secure websites are among the top 100,000 in Alexa. Therefore, if a domain is not detected in Alexa, then it indicates that the relevant website is suspicious.

- Page rank

  The page rank value is in the range of 0 and 1. The purpose of this value is to show the truth of how important a page is on the Internet. A higher value indicates the greater importance of that page. According to statistics, about 95% of fraudulent websites do not have page ranks. In addition, the rest of the scam websites have a page rank lower than 0.2.

- Google Index

  This feature indicates whether a website is in the Google index or not. When a website is indexed in Google, then the Google search engine can display it. Most scam websites are not indexed by Google because they are only available for a short period of time.

- Number of links to the page

  The number of links that point to a web page indicates how secure it is, even if some links point to the same domain. In the reviewed database, it was found that there is no external link

to 98% of fraudulent websites. Secure websites, on the other hand, have at least two external links.

3.1.5. Features based on statistical reports

Some companies, such as PhishTank and StopBadware, publish numerous statistical reports on phishing websites over time. Some of these reports are monthly and some are quarterly. In the database used in this article, 10 domains and 10 IP_s published by PhishTank and 50 IP addresses published by StopBadware from January 2010 to November 2012 have been studied to identify phishing websites.

3.2. Evaluation criteria

To evaluate the performance of the proposed system in this article, the 10-fold cross validation method will be employed as only one training and one test cannot be reliable. In 10-fold cross validation, the X database is randomly divided into 10 non-overlapping sections of equal size as X_i ,$i = 1, 2, \ldots, 10$. To produce each pair of training and test data, one of the 10 sections is used for test and the other 9 sections for training. Repeat this operation 10 times. In this way, 10 pairs are obtained as follows:

$$
\begin{array}{lll}
V_1 = X_1 & , & T_1 = X_2 \cup X_3 \cup \cdots \cup X_{10} \\
V_2 = X_2 & , & T_2 = X_1 \cup X_3 \cup \cdots \cup X_{10} \\
. & , & . \\
. & , & . \\
. & , & . \\
. & , & . \\
V_{10} = X_{10} & , & T_{10} = X_1 \cup X_2 \cup \cdots \cup X_9
\end{array}
$$

In this article, the 10-fold cross validation strategy was used 5 times for the test step and the results, and all the results related to the machine learning step are based on the average of these 5 tests.

The following criteria are usually used to evaluate the performance of classifying websites into two types of phishing and safe, which are obtained from the confusion matrix:

Table 3:

Predicted class			
Actual class		Class = phishing	Class = safe
	Class = phishing	a (TP)	d (FN)
	Class = safe	b (FP)	c (TN)

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$FNR = \frac{FN}{FN + TP} \quad (2)$$

$$FPR = \frac{FP}{FP + TN} \quad (3)$$

$$sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$specificty = 1 - FPR = \frac{TN}{TN + FP} \quad (5)$$

In the above relationships, TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative, respectively. Accuracy indicates what percent of websites are properly classified using the relevant classifier. In addition, FPR and FNR represent false positive error rates and false negative error rates. In fact, these two criteria represent system error. FP indicates what percentage of safe websites are incorrectly classified as phishing. On the other hand, a FNR also indicates what percentage of phishing websites are incorrectly classified as safe by the category.

In the real world, the costs of FPR and FNR are not the same, and FNR is more important than FPR. Because if a fraudulent website is mistakenly identified as safe, then an irreparable error will occur to the user. This mistake can cost the user information abuse. But if we introduce a safe website as a scammer, the cost of this incorrect classification can be compensated because it is only considered as a false alarm. Because of the above problem, we used a criterion called F-measure as the main criterion. This criterion is defined according to the following equation and is in fact a combination of precision and recall.

$$recall = \frac{TP}{FN + TP} \quad (6)$$

$$precision = \frac{TP}{FP + TP} \quad (7)$$

$$F - Measure = \frac{2 * precision * recall}{precision + recall} \quad (8)$$

3.3. Performance evaluation

In this section, we evaluate the performance of the proposed method in comparison with other techniques. As can be seen in Table 3, the proposed method achieves a higher accuracy in terms of F-criterion than other methods. In fact, we managed to create a classifier by combining the three AdaBoost, Bagging and Rotation Forest classifiers, which have a higher discriminability power than each of them.

In the last two rows of Table 4, we used two simple classifiers called Naïve Bayes and the decision tree. As can be seen, the accuracy of these classifiers is also above 90%. Such a phenomenon reflects the fact that the database under study was not a difficult database. In fact, the extracted features were so good that they took the issue of classification out of a complex state. The existence of such distinguishing features is the most important advantage for any classifier.

It should be noted that in the approaches tested in this paper, the classification parameters are considered by default in WEKA software.

One of the major challenges in identifying phishing websites is that scammers use a variety of methods to steal information, while existing research datasets are very old and somewhat customized. In general, there are many websites that do phishing, but it is not possible to identify

Table 4: Comparison of different classifiers on the dataset used

Classification method	Accuracy (%)	TP rate (%)	FP rate (%)	Precision (%)	Recall (%)	F-measure (%)
AdaBoost.M1	92.5	91.6	6.6	91.6	91.6	91.6
Bagging	96%	94.6	2.5	96.8	94.6	95.7
Rotation forest	96.3	95.7	3.0	96.2	95.7	95.9
Proposed method	96.7	95.6	2.3	97.1	95.6	96.3
Naïve Bayse	92.9	90.4	5.0	93.6	90.4	91.9
Decision tree	95.8	94.2	2.8	96.4	94.2	95.3

them by considering simple (though effective) features. Most of the methods in the articles focus on the classification method and how to use machine learning methods instead of feature extraction, while the need for serious studies on the extraction of effective features is strongly felt. Therefore, researchers are suggested to focus more on extracting features from phishing websites.

One of the principles on which ensemble classifiers are based is diversity. One of the ways to create diversity is to create diversity at the feature level. It takes a lot of features to create diversity at the feature level. In fact, increasing the number of useful features can increase the ability of ensemble classifiers. Unfortunately, the standard database with the most features is the same database used in this article, which contains only 30 features. Therefore, building a standard database with a large number of features is very important in this area and its need is strongly felt.

4. Conclusion

In this article, a new method was used to identify phishing websites. In the method proposed in this paper, a special combination of ensemble classifiers was used through combining three hybrid classifiers, called AdaBoost.M1, Bagging, and Rotation Forest using a stack generalization strategy. In this strategy, a separate classifier is needed to teach how classifiers vote. In this paper, a simple Bayesian classifier was used as a combiner. In this regard, for each sample in the training set, the outputs generated by the trained classifiers were retained and used as new features to build the desired dataset for the combiner. Thus, the 3 outputs generated from the basic classifiers (Bagging, AdaBoost, and rottaion forest) and the actual sample label in the training set formed a new feature vector. The training dataset for the ensemble classifier was a 4-element binary vector. Such a combiner learns how basic classifiers vote and increases classification efficiency.

References

[1] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, (2001).
[2] K. Dunham, *Mobile Malware Attacks and Defense*, (2009). Retrieved from: http://www.sciencedirect.com/science/book/9781597492980.
[3] Gartner. (2016). (Gartner) Retrieved from: http://www.gartner.com.
[4] H. Ghayoumi Zadeh, A. Montazeri, I. Abaspur Kazerouni and J. Haddadnia, *Clustering and screening for breast cancer on thermal images using a combination of SOM and MLP*, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 5(1) (2017) 68-76.
[5] M. Ghane, AR. Nejad, M. Blanke, Z. Gao and T. Moan, *Statistical fault diagnosis of wind turbine drivetrain applied to a 5MW floating wind turbine*, Journal of Physics: Conference Series 753 (5) (2017).
[6] M. Ghane and MJ. Tarokh, *Multi-objective design of fuzzy logic controller in supply chain*, Journal of Industrial Engineering International 8 (1), 1-8.
[7] M. Ghane, M. Zarvandi and MR. Yousefi, *attenuating bullwhip effect using robust-intelligent controller*, 2010 5th IEEE International Conference Intelligent Systems, (2010) 309-314.

[8] M. Khonji, Y. Iraqi and A. Jones, *Phishing Detection: A Literature Survey*, Ieee Communications Surveys & Tutorials, 15(4) (2013). Retrieved from: doi:10.1109/SURV.2013.032213.00009.
[9] E. Koozegar, M. Soryani and I. Domingues, *A New Local Adaptive Mass Detection Algorithm in Mammograms*, BIOSIGNALS. 2013.
[10] E. Kozegar, et al, *Computer aided detection in automated 3-D breast ultrasound images: a survey*, Artificial Intelligence Review (2019) 1-23.
[11] P. Likarish, D. Dunbar and T. E. Hansen, *B-apt: Bayesian antiphishing toolbar. IEEE International Conference on Communications*, (2008) 1745 –1749. Retrieved from: doi:10.1109/ICC.2008.335.
[12] G. Liu, B. Qiu and L. Wenyin, *Automatic detection of phishing target from phishing webpage*, (2010) 4153 –4156. Retrieved from: doi:10.1109/ICPR.2010.1010.
[13] O. Rahmani Seryasat and J. Haddadnia. *Evaluation of a new ensemble learning framework for mass classification in mammograms*, Clinical breast cancer 18.3 (2018) e407-e420.
[14] O. Rahmani Seryasat, J. Haddadnia and H. Ghayoumi-Zadeh, *A new method to classify breast cancer tumors and their fractionation*, Ciência e Natura, 37(4) (2015) 51-57.
[15] O. Rahmani Seryasat, J Haddadnia and H. Ghayoumi Zadeh, *Assessment of a Novel Computer Aided Mass Diagnosis System in Mammograms*, Iranian Journal of Breast Disease 9 (3) (2016) 31-41.
[16] O. Rahmani Seryasat and J. Haddadnia. *Assessment of a novel computer aided mass diagnosis system in mammograms*, Biomedical Research 28 (7) (2017).
[17] O. Rahmani Seryasat, I. Kor and H. Ghayoumi Zadeh, *Predicting the number of comments on Facebook posts using an ensemble regression model*, International Journal of Nonlinear Analysis and Applications, 12 (2021) 49-62.
[18] M. Rami, T.L. McCluskey and A. Thabtah Fadi, *Intelligent Rule based Phishing Websites Classification*, IET Information Security, 8 (2014).
[19] S.M. Sheikholeslam Noori, M. Taeibi Rahni and S.A. Shams Taleghani, *Multiple-relaxation time color-gradient lattice Boltzmann model for simulating contact angle in two-phase flows with high density ratio*, European Physical Journal Plus, 134(8) (2019) 399.
[20] A. Salmasi, A. Shadaram and A.S. Taleghani, *Effect of plasma actuator placement on the airfoil efficiency at poststall angles of attack*, IEEE Transactions on Plasma Science, 41(10) (2013) 3079–3085.
[21] Symantec, *Internet Security Threat Report*, (2014). Retrieved from: https://www.symantec.com/security-center/threat-report.
[22] A.S. Taleghani, A. Shadaram, M. Mirzaei, S. Abdolahipour, *Parametric study of a plasma actuator at unsteady actuation by measurements of the induced flow velocity for flow control*, Journal of the Brazilian Society of Mechanical Sciences and Engineering, 40(4) (2018) 173.
[23] C. Whittaker, B. Ryner and M. Nazif, *Large-scale automatic classification of phishing pages*, 10 (2010). Retrieved from http://www.internetsociety.org/sites/default/files/whit.pdf.
[24] I. Zare, A. Ghafarpour, H. Ghayoumi Zadeh, J. Haddadnia and S.M. Mostafavi Isfahani, *Evaluating the thermal imaging system in detecting certain types of breast tissue masses*, (2016).
[25] H. Zhang, G. Liu, T. Chow and W. Liu, *Textual and visual contentbased anti-phishing: A bayesian approach*, IEEE Transactions on Neural Networks, 22 (2011) 1532 –1546. Retrieved from: doi:10.1109/TNN.2011.2161999.