International Journal of Computer Applications (0975 - 8887)
Volume 122 - No.20, July 2015

On Determining the Most Effective Subset of Features for Detecting Phishing Websites

Doaa Hassan
Computers and Systems Department
National Telecommunication Institute
Cairo- Egypt

ABSTRACT

Phishing websites are a form of mimicking the legitimate ones for the purpose of stealing user 's confidential information such as usernames, passwords and credit card information. Recently machine learning and data mining techniques have been a promising approach for detection of phishing websites by distinguishing between phishing and legitimate ones. The detection process in this approach is preceded by extracting various features from a website dataset to train the classifier to correctly identify phishing sites. However, not all extracted features are effective in classification or equivalent in their contribution to its performance. In this paper, we investigate the effect of feature selection on the performance of classification for predicting phishing sites. We evaluate various machine learning algorithms using a number of feature subsets selected from an extracted feature set by various feature selection techniques in order to determine the most effective subset of features that results in best classification performance. Empirical results shows that using our new proposed methodology for selecting features by removing redundant ones that equally contribute to the classification accuracy, the decision tree classifier achieves the best performance with an overall accuracy of 95.40%, false positive rate (FPR) of 0.046 and false negative rate (FNR) of 0.065.

General Terms

Security, Algorithms

Keywords

phishing websites detection, machine learning, classification, feature selection

1. INTRODUCTION

Phishing is a form of identity theft that is usually made through emails or website in order to gain authorized access to user's private information. Phishing websites are bogus sites, where a phishing attacker attracts victims to a spoofed website similar to a legitimate one, a so-called "phishing site". Once victims access the phishing site, then the attacker attempts to convince them to send their private information such as usernames, passwords and credit card resulting in stealing their information that might cost them over a billion dollars each year.

Different research have been proposed to train some data ming and machine learning classifiers to detect phishing sites using some characteristic features of webpages [28, 8, 29]. These features are commonly divided into two categories: URL-based [16, 5, 6] features and content-based features [30, 27]. The former, refers to the URL patterns of phishing sites that distinguish them from benign ones. The later refers to the features of page's content that are used to identify phishing sites. All features from both categories serve as an inputs to machine learning classification techniques, which are then trained to identify phishing sites. However, poor performance of classifiers results from assuming that all features have an equal significance and impact on classification accuracy. Therefore, a well known approach for mitigating this problem is to use feature selection [7, 22, 24, 10, 23] in order to determine the most powerful subset of features that results in best classification accuracy. The work in this paper extends this direction by investigating various feature selections methods for determining the most effective subset of features used by classifiers to distinguish phishing websites from legitimate ones.

Given a dataset that consists of a collection of phishing sites as well as legitimate ones with some extracted features, how can we determine the most effective subset of features that has a high impact on the classification accuracy?

The work presented in this paper tries to answer this question by applying different feature selection mechanisms on a websites dataset to obtain the most effective subset of feature used by classifier to identify phishing sites. We have generated three supervised classifiers including: the Decision Tree [17, 1], Naive Bayes [15] and Suppoert vector Machine (SVM) [26] and evaluate them on a publicly available website dataset with many extracted features. Next we evaluate the three generated classifiers with various feature selection techniques including: feature selection by category, our proposed methodology for feature selection by removing redundant features with an equivalent contribution to the classification accuracy, feature selection by wrapper method [12] and feature selection by filter method [20]. Finally we compare the classification results we obtained using various feature selection techniques. Moreover we compare these results with the ones obtained using all extracted features (i.e., with no feature selection) to determine the most effective subset of feature that result in best classification performance. Our experimental results shows that the decision tree achieves the best classification performance with a feature subset selected by removing redundant features that equally contributes to classification

accuracy with an overall accuracy of 95.40%, FPR of 0.046 and FNR of 0.065.
The structure of this paper is organized as follows: in the next section, we present the related work. In Section 3 we introduce our methodology for determining the most effective subset of features with the highest impact on classification performance for detecting phishing sites. In Section 4, we present the experimental environment, focusing on describing the dataset, the extracted features and the various feature selection approaches used in the paper. In Section 5, we evaluate the performance of classifiers and state some remakes. Finally we conclude in Section 6 and investigate directions for future work.

2. RELATED WORK

There are some research attempts in the literature that investigated the effect of feature selection on improving the accuracy of classification techniques [11, 10, 14].
As for phishing detection, A. Bergholz et el. [4] presented an approach for improving of learning models for detecting phishing emails by feature selection. A subset of features is selected by a wrapper method in which the so-called best-first search algorithm systematically adds and subtracts features to a current subset using the classifier itself as part of the evaluation function.
R B. Basnet et el. [3] applied two feature selection methods: the correlation based feature selection (CFS) and feature selection with wrapper type on a real world website dataset with 177 features. They studied the effect of the two feature selection techniques on improving the performance of various classifiers for predicting phishing websites. They found that the wrapper method improves the classification results compared to CFS technique.
W. Chu et el. [6] presented a machine learning based phishing detector using only lexical and domain features, which are available even when the phishing webpages are inaccessible (since many phishing Webpages had a very short life span). They investigated the effectiveness of each feature, and each group of features on classification accuracy. Then they applied the sequential forward selection method and the plus-m-minus-r algorithm [13] to select an optimal set of features used by their phishing detector to achieve the best detection rate.
Mohammed et el. [19] presented an automatic approach for extracting features relative to phishing sites from a dataset of webpages. Though they measured the feature significance in detecting phishing by computing its frequency in the collected dataset, they did not determine the most effective subset of features to be used by learning classifiers to achieve the best detection rate of phishing websites. Our work extends their work by evaluating the performance of various supervised machine learning classifiers on their dataset. Moreover, we apply various feature selection techniques to their extracted features set in order to determine the most effective subset of features that results in the best performance of the learning classifiers for predicting the phishing sites.

3. PROPOSED METHODOLOGY

Our approach uses machine learning models for detecting phishing websites based on some extracted features from a websites dataset. However, some extracted features are redundant or have the same effect on classifier accuracy for predicting phishing sites. To this end, we aim to determine the most effective subset of features among the extracted ones from a publicly available phishing websites dataset that leads to an optimum classification performance. We achieve this by applying various feature selection techniques to remove the redundant and irrelevant features from the extracted features set. Then we create n subsets of selected features, where n represents the number of feature selection methods that we used (i.e., each subset of feature is created by one feature selection method). Later we evaluate various supervised machine learning classifiers with each features subset and pickup the most effective subset of features in predicting phishing websites that results in the best classification performance. The basic architecture of our approach for evaluation using feature selection is illustrated in Figure 1.

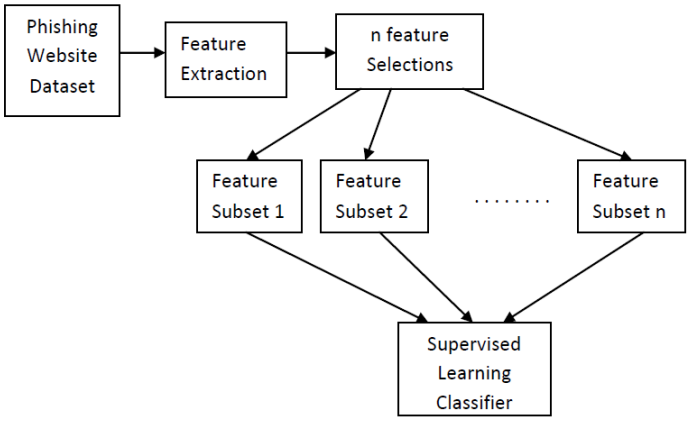

Fig. 1. **A framework of the proposed methodology for evaluation with various feature selection methods.**

4. EXPERIMENTAL ENVIRONMENT

We have run the experiments on a laptop machine with 2.3 GHz Intel processors core (TM)i3-2350M with 2 GB memory Rams. To perform feature selection and classification, we used Waikato Environment for Knowledge Analysis (WEKA), a free machine learning software tool [9].

4.1 Dataset

We use the Phishing Websites dataset available at Machine Learning Repository [2]. The dataset consists of 2456 collected Websites samples. 1362 of them are labeled as phising sites while the remaining 1094 samples are labeled as legitimate. The choice of this dataset is due to its richness in extracted features from various categories (it has 30 features divided into four groups) as we will describe in the next subsection.

4.2 Extracted Features

The extracted features from the phishing website dataset are mainly classified into four categorirs: Address Bar based Features, Abnormal Based Features, HTML and JavaScript based Features and Domain based Features. All extracted features and their category are listed in Table 1. Following we describe briefly each category and we refer to [19, 18] for more details about the extracted features.

(a) **Address Bar based Features** this refers to all features related to the address bar that shows the current URL of the examined website. An example of those features are the IP address of the domain name in the URL, URL length, URL with @ symbol, the existence of − in the domain name part of the URL, the usage of https and issuer of the website, the expiration of the

domain, open ports, Favicon (i.e.,the graphic image (icon) associated with a specific webpage) and the existence of HTTPS Token in the Domain Part of the URL.

(b) **Abnormal Based Features** this refers to all features resulting from capturing the abnormal behaviours demonstrated in website. An example of those features are examining whether a webpage contains external objects such as images [1], if the $< a >$ tags and the website have different domain names [2], if $< Meta >$, $< Script >$ and $< Link >$ tags linked to the same webpage, examining if the Server Form Handler (SFH) is an empty string or blank, if the mail()" or "mailto:" functions are used in the source code of the webpage to submit user's information to phisher's personal email and if the host name is included in the URL of the examined website.

(c) **HTML and JavaScript based Features** this refers to all features related HTML and Javascript source code of the webpages included in the examined website. An example of those features are how many times a website has been redirected, if a fake URL in the status bar is shown to the users[3], if the right click function is disabled to prevent the users from viewing and saving the webpage source code, if website asks user to submit her/his personal information through a pop-up window, and if the IFram HTML tag is used to display an additional webpage into the currently one shown.

(d) **Domain based Features** this refers to all features of the domain part in the URL of the examined website. An example of those features are checking if the websites live for a short period of time, if no DNS record (mapping between IP address and the domain associated with) exists for the domain, popularity of the website, the page rank, if the wesite is indexed by Google, number of links pointing to the webpage, and if the website is reported as a phishing site by several parties that track phishing such as phishTank and StopBadware.

Table 1. Extracted features from Phishing Website dataset described in [2]

No.	Feature	Category
1	Using the IP Address	Address Bar based Features
2	URL-Length	
3	Shortining-Service	
4	having-At-Symbol	
5	double-slash-redirecting	
6	Prefix-Suffix	
7	having-Sub-Domain	
8	SSLfinal-State	
9	Domain-registration-length	
10	Favicon	
11	port	
12	HTTPS-token	
13	Request-URL	Abnormal Based Features
14	URL-of-Anchor	
15	Links-in-tags	
16	SFH	
17	Submitting-to-email	
18	Abnormal-URL	
19	Redirect	HTML and JavaScript based Features
20	on-mouseover	
21	RightClick	
22	popUpWidnow	
23	Iframe	
24	age-of-domain	Domain based Features
25	DNSRecord	
26	web-traffic	
27	Page-Rank	
28	Google-Index	
29	Links-pointing-to-page	
30	Statistical-report	

4.3 Selecting Features

The feature selection phase follows the feature extraction one. We apply the following feature selection methods on experimental phishing website dataset in order to remove the ineffective features from the extracted features set that result in decreasing the performance of classification:

(1) **feature selection by feature category** we create four subsets of features from the extracted feature set. Each of them is equivalent to one of the feature categories illustrated in Table 1 with the same number of features elements.

(2) **feature selection by omitting redundant features** this is a new methodology we propose for feature selection. The first step in our methodology is to perform classification using each individual feature in the extracted feature set in order to determine the contribution of each feature to the classification accuracy. Next we determine the features with an equivalent contribution to the classification accuracy. We call them redundant features to distinguish them from non-redundent ones that differ in their contribution to the classification accuracy. Then we apply the following procedure for feature selection:
   —let us assume that we have a subset of redundant features with n elements and a subset of nonredundent features with m elements. Then by iterating over all the elements of redundant feature subset, choosing one of them each time plus all elements of non-redundant features subset then we create new n subsets of features, each of them with (m + 1) elements.
   —evaluate the classifier using each new subset of features created in the pervious step with (m + 1) elements.
   —pickup the subset of features from the previous step that results in the best classification performance. This is considered the subset of feature created by omitting redundant features.

(3) **feature selection by wrapper method** this is a common feature selection method in machine learning literature [25] that uses the learning algorithm itself to evaluate the usefulness of features. The process starts first by creating all possible subsets from the feature vector using different search techniques such as breadth first search, depth first search, random search, or a hybrid search. Then a classifier is induced from the features in each subset. Finally, the subset of features that leads to best classification performance is considered.

(4) **feature selection by filter method** this is another common feature selection method in machine learning literature, where all extracted features from the dataset is ranked according to weights assigned by ranker algorithms [21] based on the general characteristics of data. Then the optimum number of selected features is defined by omitting the features that have lower ranks one at time and test the predictive accuracy of the classifier.

[1]This is commonly known as request URL feature
[2]This is commonly known as request URL of Anchor feature
[3]This is commonly known as onMouseOver feature

5. PERFORMANCE EVALUATION

We feed our experimental phishing website dataset to three common supervised machine learning algorithms in order to build the classification models that detects the phishing websites. We evaluate their performance for learning phishing websites by classifying each website in the dataset as phishing or legitmate. The three generated supervised classifiers are: Decision Tree [17, 1], Naive Bayes [15] and Support Vector Machine (SVM)[26]. More precisely the J48, NaiveBayes, SMO implementations respectively in Weka. Those classifiers are from different classifiers families in Weka: J48 belongs to trees classifiers, NaiveBayes belongs to bayes classifiers, while SMO belongs to functions classifiers. We have chosen classifiers from different categories in order to build different classifications models for detecting phishing sites and compare their performance. All the generated classifiers were tested with their default parameter values using 10-fold cross-validation, a standard approach for evaluating the performance of the classifiers (where 90% of the dataset is randomly selected for training and the remaining 10% is kept for testing). The performance of each classifier is evaluated by the following metrics: the overall accuracy expressed by the number of correct predictions of phishing and legitmate sites from all predictions made by the classifier, the false positive rate (FPR) expressed by the percentage of misclassified legtimate websites instances as phishing ones, and the false negative rate expressed by the percentage of misclassified phishing websites instances as legitimate ones.

5.1 Evaluation without Feature Selection

Table 2 shows the classification results of the three generated classifiers using all extracted features from the dataset.

Table 2. The performance results of the three generated classifiers

Classifier	Accuracy	FP rate	FN rate
Decision Tree	94.99%	0.055	0.044
Naive Bayes	94.055%	0.053	0.068
Support Vector Machine (SVM)	94.63%	0.048	0.060

As can be seen clearly from the table, the values of overall accuracy, FPR and FNR for all classifiers are closed to each other. The decision tree classier slightly outperforms the other two classifiers in overall accuracy and FNR as it achieves 94.99% and 0.044 respectively. Next SVM classifier achieves an accuracy of 94.62% and FNR of 0.06 , while Naive Bayes classifier achieves an accuracy of 94.06% and FNR of 0.068. On the other hand, the SVM outperforms the other two classifiers in FPR as it achieves 0.048. Next Naive Bayes achieves FPR of 0.053 then decision tree achieves FPR of 0.055.

5.2 Evaluation with Feature Selected by their Category

We performed detection of phishing websites by classification using the group of features for each feature category shown in Table 1. Figure 2 and Table 3 show the resulting accuracy, FPR, and FNR respectively of the generated classifiers using each group of features for each feature category. Clearly seen from the figure that the classification using the subset of features for address bar based features category leads to the best accuracy. It has an overall accuracy of 91.49% with decision tree, 90.64% with Naive bayes, and 89.58% with SVM. The abnormal based features and domain based features categories are the next two most effective subsets of features on classification accuracy respectively. They lead to an overall accuracy of 86.20% and 80.54% with decision tree, 85.75% and 80.91% with Naive Bayes, and 84.04% and 79.40% with SVM. Finally, the group of features in HTML and JavaScript based features category is the least effective subset of features on classification accuracy. It leads to an overall accuracy of 56.92% with decision tree, 53.83% with Naibe Bayes, and 54.72% with SVM.

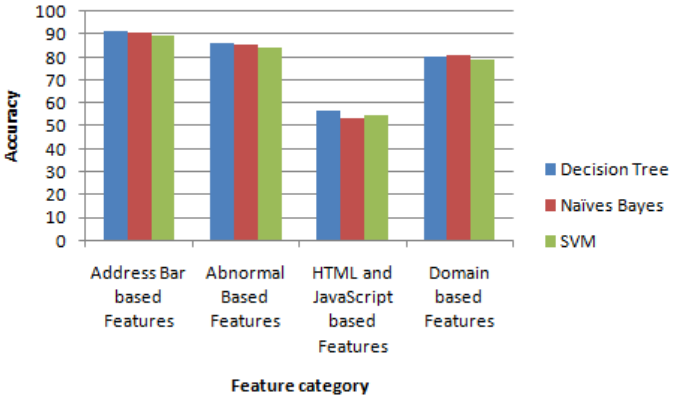

Fig. 2. The accuracy of classifiers for predicting phishing sites using feature category.

5.3 Evaluation with Omitting Redundant Features

The first step in our proposed new methodology for evaluating classifiers using a subset of features selected by removing redundant ones is to perform classification using each individual feature to determine its contribution to the classification accuracy and hence its effectiveness in identifying phishing websites. Figure 3 shows the resulting accuracy of each classifier for predicting phishing sites using each individual feature in the dataset. From the figure, the following remarks can be drawn:

—the SSLfinal-State is the most effective feature among all 30 features on classification accuracy as it leads to an overall accuracy of 88.68% for all generated classifiers. URL-of-Anchor and web-traffic are the next most effective features leading to an overall accuracy of 84.04% and 78.01% respectively for all generated classifiers.

—The on-mouseover and port features are the least two effective features among all 30 extracted features with lowest impact on classification accuracy. The former leads to the lowest overall accuracy of 54.89% with decision tree classifier, while the later leads to the lowest overall accuracy of 54.60% with Naive Bayes and SVM classifiers.

—The having-IP-Address, Shortining-Service, having-At-Symbol, double-slash-redirecting, Favicon, HTTPS-token, SFH, Submitting-to-email, Abnormal-URL, Redirect, RightClick, popUpWidnow, Iframe, Links-pointing-to-page, Statistical-report features are redundant features that equally contribute to the classification accuracy as each of them results in an overall accuracy of 55.46% for all classifiers.

—The Decision Tree outperforms SVM and Naive Bays in overall accuracy using either URL-length or port features. The former achieves an overall accuracy of 55.46% in case of URL-length and port features, while SVM and Naive Bayes achieve

Table 3. The resulting FPR and FNR for each classifier using a feature subset selected according to feature category.

Feature category	Decision Tree		Naive Bayes		SVM	
	FPR	FNR	FPR	FNR	FPR	FNR
Address Bar based Features	0.08	0.09	0.10	0.09	0.07	0.15
Abnormal Based Features	0.05	0.28	0.06	0.25	0.01	0.35
HTML and JavaScript based Features	0.012	0.95	0.08	0.93	0.07	0.93
Domain based Features	0.17	0.23	0.16	0.23	0.17	0.26

an overall accuracy of 55.62% in case of URL-length feature and 54.60% in case of port feature.

Using the results of evaluation with individual features obtained above, we can determine the features with an equivalent contribution to the classification accuracy. Hence we can apply the procedure presented in Section 4.3 for feature selection by omitting redundant ones to select the most effective subset of features that results in best classification results.
Table 4 shows the subsets of features generated by the procedure in 4.3 that lead to the best performance for each classifier.
The feature combination that achieves the best overall accuracy of 95.40% for decision tree classifier comprises of the following 14 features (listed by their index in Table 1): 6, 7, 8, 9, 10, 13, 14, 15, 20, 24, 25, 26, 27, 28.
The feature combination that achieves the best overall accuracy of 94.055% for Naive Bayes classifier comprises of the following 16 features: 2, 3, 6, 7, 8, 9, 11, 13, 14, 15, 20, 24, 25, 26, 27, 28.
Two feature combinations achieve the best overall accuracy of 94.87% for SVM classifier. The first comprises of the following 16 features: 2,6,7,8,9,11,13,14,15,19,20,24,25,26,27,28. The second comprises of the same 16 features in the first combination except replacing the Redirect feature with popUpWidnow.

5.4 Evaluation with Features Selected with Wrapper Method

Table 5 shows the most effective subset of features selected by wrapper method that results in the best performance for each classifier.
The feature combination that achieves the best overall accuracy of 94.67 % for decision tree classifier comprises of the following 12 features (listed by their index in Table 1): 2, 6, 7, 8, 9, 11, 14, 15, 24, 25, 26, 30.
The feature combination that achieves the best overall accuracy of 93.93% for Naive Bayes classifier compromises of the following 10 features: 2, 6, 7, 8, 9, 14, 16, 26, 28, 29.
The feature combination that achieves the best overall accuracy of 94.54% for SVM classifier compromises of the following 18 features: 1, 2, 3, 5, 6, 7, 8, 10, 13, 14, 16, 19, 21, 22, 26, 27, 28, 30.

5.5 Evaluation with Features Selected with Filter Method

We have ranked all extracted features and selected the first n features using the filter method. We choose n to be equal to the number of features that leads to the best accuracy for each classifier using our methodology for selecting features by removing redundant ones. This will help us to compare both feature selections methods and determine which one outperforms the other. By selecting the first 14 features using the filter method and use them to train j48 classifier we got an overall accuracy of 94.75 %. We also get the same accuracy of j48 if we select the first 16 features by the same method and use them to train the classifier. We have also used those first 16 features to train Naive bayes and SVM classifiers, we got an overall accuracy of 94.096% and 94.75% respectively. Table 6 shows the first 16 effective features selected by filter method and the corresponding performance for each classifier.

5.6 Evaluation Remarks

By analyzing the classification results we obtained using different feature selection techniques and comparing them to those obtained without feature selection, the following remarks can be concluded:

—our new proposed methodology for selecting features by removing redundant ones achieves the best performance results with the decision tree classifier with an overall accuracy of 95.40%, FPR of 0.046 and FNR of 0.065.

—though there are various features that contribute equally to the classification accuracy, selecting one of them each time plus the remaining non-redundant features to train the classifier will not result in the the same performance when replacing the selected redundant feature with another redundant one and re-perform the classification. This is due to the variations in FPR and FNR when evaluating the classifier using each of the redundant features individually even if the resulting accuracy is the same.

—our approach for selecting features by removing redundant ones is different form forward/backward approaches for feature selection implemented in wrapper method [13] as the later do not remove redundant features and they only aim to add features that improve the prediction in case of forward approach or remove features that decrease the prediction in case of backward approach.

6. CONCLUSION AND FUTURE WORK

This paper applies various feature selection techniques on a phishing website dataset in order to select the most effective subset of features from the extracted features set that results in an optimum classification performance for predicating phishing sites. We run our experimental classifiers with four feature selection methods including: feature selection by category, our proposed methodology for feature selection by omitting redundant features, feature selection by wrapper and filter methods respectively. Our experimental results shows that the decision tree classifiers achieves the best performance with a feature subset selected by omitting redundant features.
As a future work, we are looking forward to applying our proposed methodology for evaluation with the presented feature selection methods on other phishing websites datasets with larger size and more extracted features, then testing the performance of classification algorithms for identifying phishing sites. This might lead to discover new subsets of features with a high impact on the classification performance.

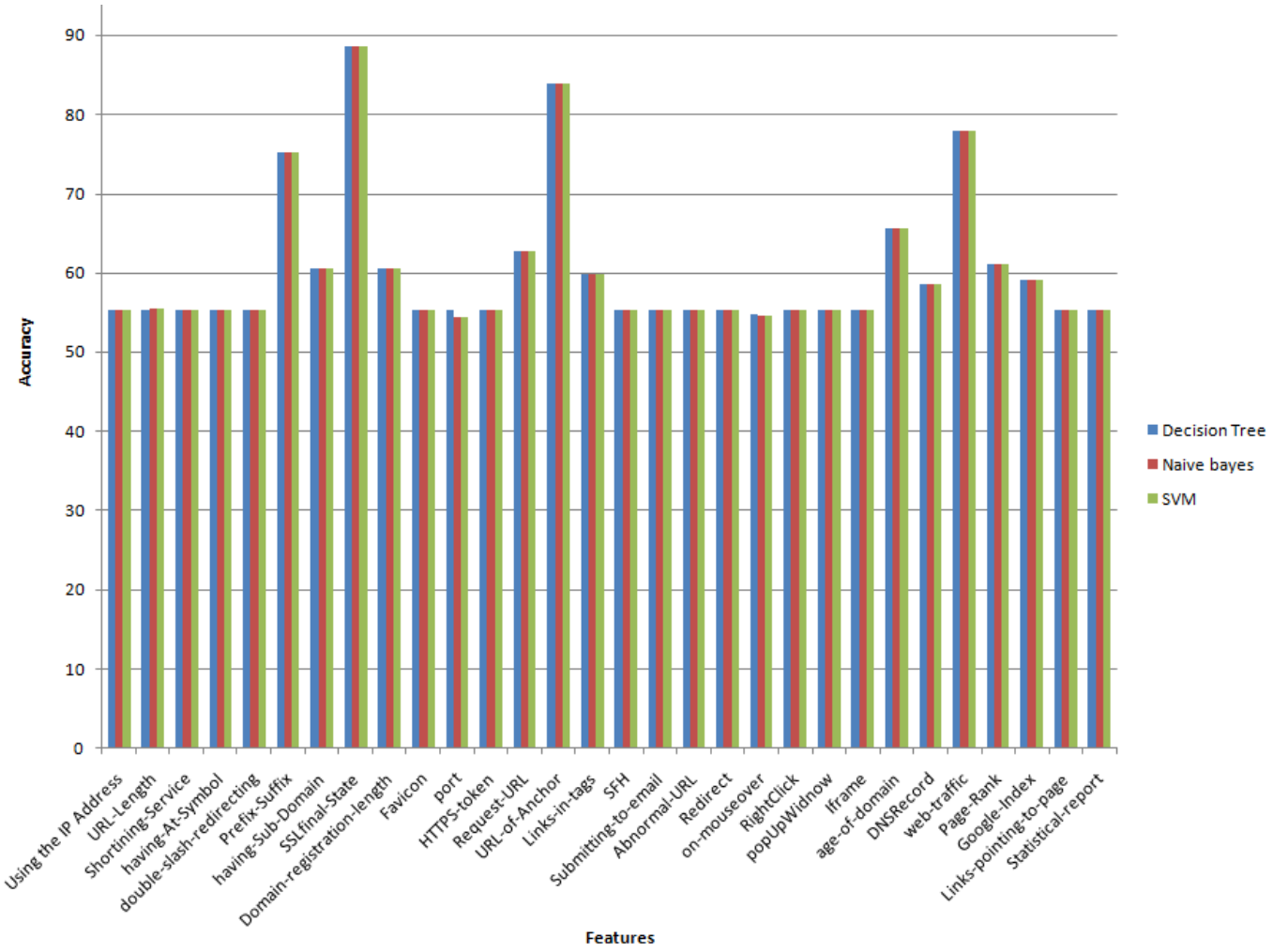

Fig. 3. **The accuracy of classifiers for predicting phishing sites using each individual feature.**

Table 4. The most effective subset of features selected by removing redundant ones and the corresponding performance for each classifier.

Classifier	Selected Features	Accuracy	FP rate	FN rate
Decision Tree	6, 7, 8, 9, 10, 13, 14, 15, 20, 24, 25, 26, 27, 28	95.40%	0.046	0.065
Naive Bayes	2, 3, 6, 7, 8, 9, 11, 13, 14, 15, 20, 24, 25, 26, 27, 28	94.055%	0.047	0.075
Support Vector Machine (SVM)	2,6,7,8,9,11,13,14,15,19,20,24,25,26,27,28 2,6,7,8,9,11,13,14,15,20,22,24,25,26,27,28	94.87%	0.049 0.048	0.054 0.055

Table 5. The most effective subset of features selected by wrapper method that leads to best performance for each classifier.

Classifier	Selected Features	Accuracy	FP rate	FN rate
Decision Tree	2, 6, 7, 8, 9, 11, 14, 15, 24, 25, 26, 30	94.67%	0.052	0.055
Naive Bayes	2, 6, 7, 8, 9, 14, 16, 26, 28, 29	93.93%	0.045	0.08
Support Vector Machine (SVM)	1, 2, 3, 5, 6, 7, 8, 10, 13, 14, 16, 19, 21, 22, 26, 27, 28, 30	94.54%	0.052	0.058

Table 6. The most effective 16 features selected by filter method that leads to best performance for each classifier.

Classifier	Selected Features	Accuracy	FP rate	FN rate
Decision Tree	8,14,6,26,27,7,24,9,13,15,25,28,29,16,2,1	94.75%	0.056	0.048
Naive Bayes	8,14,6,26,27,7,24,9,13,15,25,28,29,16,2,1	94.096%	0.051	0.069
Support Vector Machine (SVM)	8,14,6,26,27,7,24,9,13,15,25,28,29,16,2,1	94.75%	0.051	0.55

7. REFERENCES

[1] http://www.support-vector-machines.org/.

[2] https://archive.ics.uci.edu/ml/datasets/Phishing+Websites.

[3] Ram B. Basnet, Andrew H. Sung, and Quingzhong Liu. Feature selection for improved phishing detection. In *Proceedings of the 25th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems: Advanced Research in Applied Artificial Intelligence*, IEA/AIE'12, pages 252–261, 2012.

[4] Andr Bergholz, Gerhard Paa, Frank Reichartz, Siehyun Strobel, and Schlo Birlinghoven. Improved phishing detection using model-based features. In *Fifth Conference on Email and Anti-Spam, CEAS*, 2008.

[5] Aaron Blum, Brad Wardman, Thamar Solorio, and Gary Warner. Lexical feature based phishing url detection using online learning. In *Proceedings of the 3rd ACM Workshop on Artificial Intelligence and Security*, AISec '10, pages 54–60, 2010.

[6] Weibo Chu, Bin B. Zhu, Feng Xue, Xiaohong Guan, and Zhongmin Cai. Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing urls. In *Proceedings of IEEE International Conference on Communications, ICC 2013, Budapest, Hungary, June 9-13, 2013*, pages 1990–1994, 2013.

[7] M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis*, 1:131–156, 1997.

[8] Ian Fette, Norman Sadeh, and Anthony Tomasic. Learning to detect phishing emails. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 649–656, 2007.

[9] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.

[10] A. G. Janecek, W. N. Gansterer, M. A. Demel, and G. F. Ecker. On the Relationship Between Feature Selection and Classification Accuracy. In *JMLR: Workshop and Conference Proceedings 4*, pages 90–105, 2008.

[11] Esra Mahsereci Karabulut, Selma Aye zel, and Turgay briki. A comparative study on the effect of feature selection on classification accuracy. *Procedia Technology*, 1(0):323 – 327, 2012. First World Conference on Innovation and Computer Sciences (INSODE 2011).

[12] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artif. Intell.*, 97(1-2):273–324, December 1997.

[13] L. Ladha and T. Deepa. Features selection methods and algorithms. *International Journal on Computer Science and Engineering (IJCSE)*, 3(5):1787–1797, May 2011.

[14] Chang-Hwan Lee, Fernando Gutierrez, and Dejing Dou. Calculating feature weights in naive bayes with kullback-leibler measure. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining*, ICDM '11, pages 1146–1151, 2011.

[15] K. Ming Leung. Naive bayesian classifier, 2007. Department of Computer Science / Finance and Risk Engineering- Polytechnic University.

[16] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. Beyond blacklists: Learning to detect malicious web sites from suspicious urls. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 1245–1254, 2009.

[17] Tom Mitchell. *Machine Learning*, chapter Decision Tree Learning, pages 52–78. McGraw-Hil, 1997.

[18] Rami M. Mohammad, Fadi Thabtah, and Lee McCluskey. Phishing websites features, 2015. Unpublished. Available via: `http://eprints.hud.ac.uk/24330/6/RamiPhishing_Websites_Features.pdf`.

[19] Rami M. Mohammad, Fadi A. Thabtah, and Lee McCluskey. An assessment of features related to phishing websites using an automated technique. In *7th International Conference for Internet Technology and Secured Transactions, ICITST 2012, London, United Kingdom, December 10-12, 2012*, pages 492–497, 2012.

[20] Sánchez-Maro no, Noelia, Alonso-Betanzos, Amparo, and María Tombilla-Sanromán. Filter methods for feature selection: A comparative study. In *Proceedings of the 8th International Conference on Intelligent Data Engineering and Automated Learning*, IDEAL'07, pages 178–187, 2007.

[21] Jasmina Novakovic, Perica Strbac, and Dusan Bulatovic. Toward optimal feature selection using ranking methods and classification algorithms. *Yugoslav Journal of Operations Research*, 21(1):119–135, 2011.

[22] Selwyn Piramuthu. Evaluating feature selection methods for learning in data mining applications. *European Journal of Operational Research*, 156(2):483–494, 2004.

[23] M. Ramaswami and R. Bhaskaran. A study on feature selection techniques in educational data mining. *CoRR*, abs/0912.3924, 2009.

[24] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, September 2007.

[25] Luis Talavera. An evaluation of filter and wrapper methods for feature selection in categorical clustering. In *6th International Symposium on Intelligent Data Analysis*, IDA'05, 2005.

[26] Jason Weston. Support vector machine tutorial.

[27] Joshua S. White, Jeanna N. Matthews, and John L. Stacy. A method for the automated detection phishing websites through both site characteristics and image analysis, 2012.

[28] Colin Whittaker, Brian Ryner, and Marria Nazif. Large-scale automatic classification of phishing pages. In *Proceedings of the Network and Distributed System Security Symposium, NDSS 2010, San Diego, California, USA, 28th February - 3rd March 2010*, 2010.

[29] Guang Xiang, Jason I. Hong, Carolyn Penstein Rosé, and Lorrie Faith Cranor. CANTINA+: A feature-rich machine learning framework for detecting phishing web sites. *ACM Trans. Inf. Syst. Secur.*, 14(2):21, 2011.

[30] Yue Zhang, Jason I. Hong, and Lorrie Faith Cranor. Cantina: a content-based approach to detecting phishing web sites. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 639–648, 2007.