DETECTION OF PHISHING WEBSITES USING MACHINE LEARNING TECHNIQUES

[1]Mrs.Vaneeta M, [2]Pratik N N,[3]Prajwal D, [4]Pradeep K S, [5]Suhas Kakade K

[1]Associate professor, Department of Computer Science and Engineering, K S Institute of Technology
[2,3,4,5]Undergraduates, Computer Science and Engineering, K S Institute of Technology,
Bengaluru, Karnataka, India-560109, Affiliated to VTU, Belagavi

Abstract: Phishing is a website forgery with an intention to track and steal the sensitive information of online users. It is a form of identity theft, in which criminals build replicas of target websites and lure unsuspecting victims to disclose their sensitive information like passwords, PIN, etc. A huge volume of information is downloaded and uploaded constantly to the web. This gives opportunities for criminals to hack important personal information. To overcome the issues faced here,developed a phishing websites detection technique based on machine learning classifiers with a wrapper features selection method. Classification algorithms used are Artificial Neural Network, Random Forest and Support Vector Machine. Dynamic features extraction is made from the entered URL and the trained model is used for the detection of phishing URL.

Index Terms **- Phishing, Phishing attack, Wrapper Features Selection, Machine Learning Classifiers,WHOIS Protocol, Dynamic Features Selection**

I. INTRODUCTION

In recent years, the web has evolved explosively due to the availability of numerous services such as online banking, entertainment, education, and social networking. Accordingly, a huge volume of information is downloaded and uploaded constantly to the Web. This gives opportunities for criminals to hack important personal or financial information, such as usernames, passwords, account numbers and national insurance numbers. This is called a Web phishing attack, which is considered as one of the major problems in Web security.

The number of phishing attacks has been growing considerably in recent years and is considered as one of the most dangerous modern internet crimes, which may lead individuals to lose confidence in e-commerce. Consequently, it has a tremendous negative effect on online banking, e-commerce, online marketing efforts, organization's incomes, relationships with customers, and overall business operations.

The success of phishing website detection techniques mainly depends on recognizing phishing websites accurately and within an acceptable timescale. Many conventional techniques based on fixed black and white listing databases have been suggested phishing websites. However, these techniques are not efficient enough, since a new website can be launched within few seconds. Therefore, most of these techniques are not able to make an accurate decision dynamically, whether the new website is phishing or not. Hence, many new phishing websites may be classified as legitimate websites.

Here developed a phishing website detection scheme using a Wrapper feature selection technique with machine learning classifiers, to detect phishing websites with high accuracy.

Classification techniques emoployed are:

- Neural network
- Support vector machine
- Random forest.

Features considered for detecting phishing sites are grouped as follows:

- URL-Based Features
- Domain-Based Features
- Page-Based Features
- Content-Based Features

II. LITERATURE SURVEY

2.1 PhishShield: A Desktop Application to Detect Phishing Webpages through Heuristic Approach[1]

In this paper, implemented a desktop application called Phish Shield, which concentrates on URL and Website Content of phishing page. Phish Shield takes URL as input and outputs the status of URL as phishing or legitimate website. The heuristics used to detect phishing, are footer links with null value, zero links in body of html, copyright content, title content and website identity. Phish Shield is able to detect zero hour phishing attacks which blacklists unable to detect and it is faster than visual based assessment techniques that are used in detecting phishing.

2.2 Detection of Phishing Website Using Machine Learning[2]

The proposed model focuses on identifying the phishing attack based on checking phishing websites features, Blacklist and WHOIS database. According to few selected features they differentiate between legitimate and spoofed web pages. These selected features are many such as URLs, domain identity, security & encryption, source code, page style and contents, web address bar and social human factor.

2.3 Intelligent phishing url detection using association rule mining[3]

This paper focuses on discerning the significant features that discriminate between legitimate and phishing URLs. These features are then subjected to associative rule mining—a priori and predictive a priori. The rules obtained are interpreted to emphasize the

features that are more prevalent in phishing URLs. Analyzing the knowledge accessible on phishing URL and considering confidence as an indicator, the features like transport layer security, unavailability of the top level domain in the URL and keyword within the path portion of the URL were found to be sensible indicators for phishing URL. In addition to this, number of slashes in the URL, dots in the host portion of the URL and length of the URL are also the key factors for phishing URL.

2.4 Detecting Phishing Emails Using Machine Learning Techniques[4]

In this paper, various classification algorithms are discussed and compared, such as; Naïvebayes, Decision Tree (DT), Logistic Regression, Classification and Regression Trees and Sequential Mini mal Optimization (SMO). A new system was built to detect the phishing emails in an integration between the supervised and unsupervised machine learning techniques. In addition, the study compares the manual and automated feature selection groups for the Email.

2.5 Learning to Detect Phishing URLs[5]

In this paper, the study is based on the anatomy of phishing URLs that are created with the specific intent of impersonating a trusted third party to trick users into divulging personal data. Unlike previous work in this area, they only use a number of publicly available features on URL alone; in addition, compare performance of different machine learning techniques and evaluate the efficiancy of real-time application of their method. The paper proposes a heuristic-based approach to classify phishing URLs by using the information available only on URLs. They treat the problem of detecting phishing URLs as a binary classification problem with phishing URLs belong to the positive class and benign URLs belong to the negative class. They first run number of scripts to collect our phishing and benign URLs and create our data sets.The next batch of scripts then extracts a number of features by employing various publicly available resources in order to classify the instances into their corresponding classes. They then apply various machine learning algorithms to build models from training data, which is comprised of pairs of feature values and class labels.Separate set of test data are then supplied to the models, and the predicted class of the data instance is compared to the actual class of the data to compute the accuracy of the classification models.

2.6 Phishing Detection: Analysis of Visual Similarity Based Approaches[6]

Phishing website looks very similar in appearance to its corresponding legitimate website to deceive users into believing that they are browsing the correct website. Visual similarity based phishing detection techniques utilize the feature set like text content, text format, HTML tags, Cascading Style Sheet (CSS), image, and so forth, to make the decision. These approaches compare the suspicious website with the corresponding legitimate website by using various features and if the similarity is greater than the predefined threshold value then it is declared phishing.

This paper presents a comprehensive analysis of phishing attacks, their exploitation, some of the recent visual similarity-based approaches for phishing detection, and its comparative study. This survey provides a better understanding of the problem, current solution space, and scope of future research to deal with phishing attacks efficiently using visual similarity-based approaches. They broadly classify the visual similarity based approaches into HTML document object model (DOM) tree, visual features, Cascading Style Sheet (CSS) similarity, pixel based, visual perception, and hybrid approaches.

2.7 Detecting Phishing Websites via Aggregation Analysis of Page Layouts[7]

This paper proposed a learning-based mechanism to evaluate the similarity of web page layouts and identify phishing pages and define the rules to extract and create effective page layout features and develop a phishing page classifier based on two typical learning algorithms, support vector machine and decision tree. Cascading Style Sheets (CSS) is the commonly used visual layout definition of web pages.The goal is to develop methods that can detect the similarity among two page layouts by comprehensively considering layout features. The problem addressed by this paper can be formulated as follows: Taking a suspicious page and a set of benign pages as inputs, aim to extract features and apply learning algorithms to detect phishing pages comprehensively based on its layout similarity features. In this classifier, they take labelled comparison vectors as inputs (where 1 denotes that the two pages are similar and 0 denotes that the two pages are not similar). The output of the classifier is 1 (similar) or 0 (different). This involves Property Vector Generation which present the rules that quantify the CSS elements impact of a web page and combine CSS features of two pages into one comparison property vector and Classifier Building where they set the output of the classifier as a binary output, 1 or 0, and make the comparison property vectors in the dataset as inputs.

III. DESIGN AND IMPLEMENTATION

3.1 System Architecture:

The process begins with acquiring of the raw dataset from UCI and Kaggle proprietary websites. This gathered raw dataset is pre-processed to improve its reliability and usefullness for the machine learning algorithms to learn from this dataset. Data cleansing is carried out to remove junk or missing values or invalid values that could cause difficulty to digest by the ML Algorithms. Better the data more easier it is for the ML algorithms to produce better results.

In the next stage ML models are built by making use of versatile algorithms. Algorithms used are Artificial neural network, Random forest and Support vector machines. Each algorithm has its advantages and is harnessed in producing the final outcome. Each model would produce its output i.e its prediction about whether the passed data has features that could be either a phished website or a legitimate one. Next we are comparing the results produced and the most effective algorithm is chosen to be the best model. Further to enhance its performance and reliability Dynamic features extraction is performed to obtain the dataset of newly entered URL by the user and the algorithm is able to produce accurate results for newly entered data. Finally as an end result the software that makes use of the above process must be able to predict whether the URL of the website that the user wishes to access is Phishing or Legitimate.

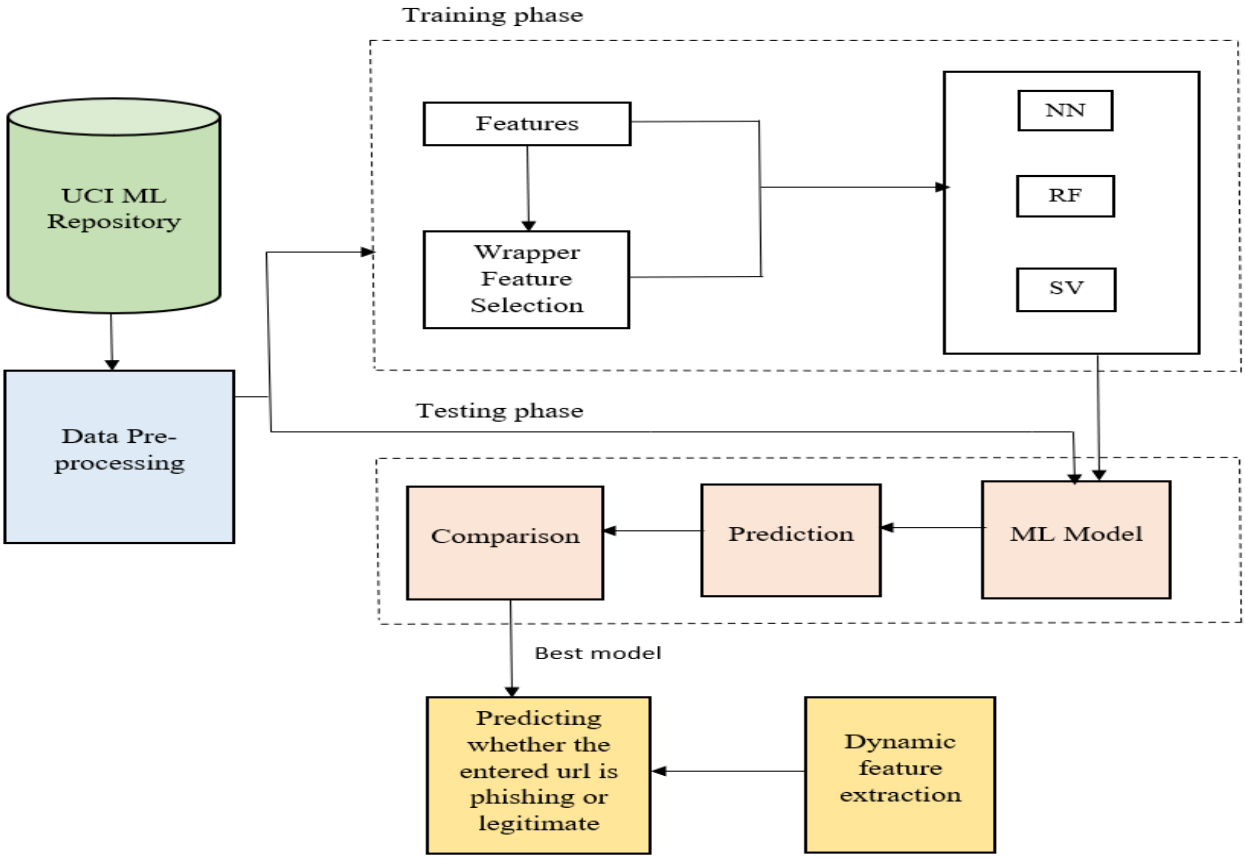

Figure 1: Architecture of the Proposed System

3.2 Modules included in the system are:

1. Data acquisition and Pre-processing
2. Feature Selection
3. Model Building and Training
4. Model Comparison and Result Analysis
5. Dynamic Feature Extraction from Entered URL
6. Detecting whether the Entered URL is Phishing or not

3.2.1 Data acquisition and Pre-processing

Machine learning needs two things to work, data (lots of it) and models. When acquiring the data, be sure to have enough features populated to train correctly our learning model. The primary data collected from the online sources remains in the raw form of statements, digits and qualitative terms. The raw data contains error, omissions and inconsistencies. It requires corrections after careful scrutinizing the completed questionnaires. A huge volume of raw data collected through field survey needs to be grouped for similar details of individual responses. Data Pre-processing is necessary because of the presence of unformatted real- world data. Initially data is collected from UCI and Kaggle proprietary websites.

3.2.2 Features Selection

The data features that use to train the machine learning models have a huge influence on the performance. Irrelevant or partially relevant features can negatively impact model performance. We use Wrapper feature selection method for selecting the features. Wrapper methods are based on greedy search algorithms as they evaluate all possible combinations of the features and select the combination that produces the best result for a specific machine learning algorithm. Here step backward selection is used. In the first step of the step backwards feature selection, one feature is removed in round-robin fashion from the feature set and the performance of the classifier is evaluated. The feature set that yields the best performance is retained. In the second step, again one feature is removed in a round-robin fashion and the performance of all the combination of features except the 2 features is evaluated. This process continues until the specified number of features remains in the dataset.

3.2.3 Model Building and Training

The process of training an ML model involves providing an ML algorithm with training data to learn from. The term ML model refers to the model that is created by the training process.

Algorithms used for training the Model:

- Neural network
- Random forest
- Support vector machine.

Random Forest: Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. Here 50 estimators are considered for the training.

Artificial Neural Network: Artificial Neural Networks (ANN) are multi-layer fully connected neural nets. They consist of an input layer, multiple hidden layers, and an output layer. Every node in one layer is connected to every other node in the next layer. We

make the network deeper by increasing the number of hidden layers. A given node takes the weighted sum of its inputs, and passes it through a non-linear activation function. This is the output of the node, which then becomes the input of another node in the next layer. The signal flows from left to right, and the final output is calculated by performing this procedure for all the nodes. Training this deep neural network means learning the weights associated with all the edges.
The input layer is the dataset and output is the classification result. And 30 epoches considered for the training.

Support Vector Machine: The Support Vector Machine Algorithm is used for classification or regression problems. In this, the data is divided into different classes by finding a particular line (hyperplane) which separates the data set into multiple classes. The Support Vector Machine Algorithm tries to find the hyperplane that maximizes the distance between the classes (known as margin maximization) as this increases the probability of classifying the data more accurately.
The Kernal function used here is radial basis kernel function(rbf) since it works on non linear seperable boundaries also.

3.2.4 Model Comparison and Result analysis

The dataset is divided into 70:30 ratio for training and testing. In testing phase the model is applied to new set of data. The train and test data are two different datasets. We used confusion matrix and accuracy score to get the performance of the model.

Confusion Matrix: A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class.

Accuracy: Accuracy is the fraction of predictions our model got right.
Accuracy=Number of correct predictions / Total number of predictions

3.2.5 Dynamic Feature Extraction from Entered URL

WHOIS Database:
The life of phishing site is very short, therefore; this DNS information may not be available after some time. If the DNS record is not available anywhere then the website is phishing. If the domain name of the suspicious webpage is not match with the WHOIS database record, then webpage considerd as phishing.

Features considered are listed below:

1. *Having an internet protocol (IP) Address*

If the IP address is used as the domain of the URL, such as http://125.98.2.142/contoh.html, it can be suspected that attempts to steal information.

2. *URL Length*

Long URLs can also be suspected of being a phishing site. If the URL length is greater than or equal to 75 characters then the URL included as a phishing site.

3. *Shortening Service:*

URL shortening is a method in which a URL is made to be shorter, which this domain ill connect to the web page that has a URL that is longer such as, http://sekolah.ini.ac.id/ URLs can be shortened to "bit.ly/21FXWl5".

4. *Having @ Symbol:*

URLs using the symbol @ will lead to the browser to ignore everything that precedes the @ symbol.

5. *Double Slash Redirecting:*

Double slash or "//" indicates that the user will be redirected to another site. The position of the use of double slash usually appears at the sixth position as written at this link http://amikom.ac.id.

6. *Prefix or Suffix:*

Rarely a legitimate URL using symbols dashboard, but phisher will add a prefix or suffix to be separated by (-) in the domain name, so the user will think to have a legitimate access to sites such as http://www.amikom-keren.com.

7. *Having Sub Domains*

The domain name may have a code for each country (cc TLD) such as "id", or for an academic educational institution "ac" and combined "ac.id" or also called two-level domain (SLD). Stages for extracting the feature of the first to do is remove the "www" in the URL and remove cc DTL if any. Then calculate the remaining dots, if the number of points is greater than one, then the URL can be classified as "suspect" because of only the subdomain. However, if the number of points greater than the two it will be categorized as a phishing because it has several subdomains, and sites categorized as legitimate if it does not have a subdomain.

8. *HTTPS (Hyper Text Transfer Protocol with Secure Sockets Layer):*

HTTPS is an essential component in a site that looks at legality.

9. *Domain Registration Length:*

Categorized as phishing sites are valid only in a short time and are used for a single year.

10. *Favicon:*

Favicon is an image used as an icon on a website, favicon also indicates the identity of the website. However, if the favicon is displayed apart in the address bar, it can be suspected that the website is a phishing website.

11. *Port:*

Port used to validate certain services such as HTTP. The use of a firewall, proxy, and Network Address Translation or NAT can perform automatic blocking and can be opened in accordance with the wishes. But if all the ports are opened, then the phisher will find loopholes and enable any desired services such as stealing information.

12. *HTTPS Token:*

In general, the https token can be added by phisher on the domain URL and has the objective to distract the user like at http://https-www-amikom-coolest-college.com/

13. *Request URL*

On the website is legal, website addresses, pictures, videos, and sounds contained on the web page with the same domain and does not take away from another domain.

14. *Abnormal URL:*

On the website is legitimate, then the identity of the website will be contained in the URL.

15. *Links in Tags:*

Given that our investigation covers all angles likely to be used in the webpage source code, we find that it is common for legitimate websites to use tags to offer metadata about the HTML document.

16. *SFH:*

SFHs that contains an empty string or "about:blank" are considered doubtful because an action should be taken upon the submitted information. In addition, if the domain name in SFHs is different from the domain name of the webpage, this reveals that the webpage is suspicious because the submitted information is rarely handled by external domains.

17. *Submitting to Email:*

The official website will generally send personal information to the server for processing. While the phisher will be sending the information to his personal email, it can be suspected by the use of scripts on the server side functions such as "mail ()" and on the client side will use the mailto().

18. *Abnormal URL:*

This feature can be extracted from WHOIS database. For a legitimate website, identity is typically part of its URL.

19. *Redirect:*

The fine line that distinguishes phishing websites from legitimate ones is how many times a website has been redirected. In our dataset, we find that legitimate websites have been redirected one time max. On the other hand, phishing websites containing this feature have been redirected at least 4 times.

20. *On_mouseover:*

Phishers may use JavaScript to show a fake URL in the status bar to users. To extract this feature, we must dig-out the webpage source code, particularly the "onMouseOver" event, and check if it makes any changes on the status bar.

21. *Right Click:*

Phishers use JavaScript to disable the right-click function, so that users cannot view and save the webpage source code. This feature is treated exactly as "Using onMouseOver to hide the Link". Nonetheless, for this feature, we will search for event "event.button==2" in the webpage source code and check if the right click is disabled.

22. *Popup Window:*

It is unusual to find a legitimate website asking users to submit their personal information through a pop-up window. On the other hand, this feature has been used in some legitimate websites and its main goal is to warn users about fraudulent activities or broadcast a welcome announcement, though no personal information was asked to be filled in through these pop-up windows.

23. *Iframe:*

IFrame is an HTML tag used to display an additional webpage into one that is currently shown. Phishers can make use of the "iframe" tag and make it invisible i.e. without frame borders. In this regard, phishers make use of the "frameBorder" attribute which causes the browser to render a visual delineation.

24. *Age of Domain:*

This feature can be extracted from WHOIS database. Most phishing websites live for a short period of time. By reviewing our dataset, we find that the minimum age of the legitimate domain is 6 months.

25. *DNS Record:*

For phishing websites, either the claimed identity is not recognized by the WHOIS database or no records founded for the hostname. If the DNS record is empty or not found then the website is classified as "Phishing", otherwise it is classified as "Legitimate".

26. *Web traffic:*

This feature measures the popularity of the website by determining the number of visitors and the number of pages they visit. However, since phishing websites live for a short period of time, they may not be recognized by the Alexa database. By reviewing our dataset, we find that in worst scenarios, legitimate websites ranked among the top 100,000. Furthermore, if the domain has no traffic or is not recognized by the Alexa database, it is classified as "Phishing". Otherwise, it is classified as "Suspicious".

27. *Page Rank:*

PageRank is a value ranging from "0" to "1". PageRank aims to measure how important a webpage is on the Internet. The greater the PageRank value the more important the webpage. In our datasets, we find that about 95% of phishing webpages have no PageRank. Moreover, we find that the remaining 5% of phishing webpages may reach a PageRank value up to "0.2".

28. *Google Index:*

This feature examines whether a website is in Google's index or not. When a site is indexed by Google, it is displayed on search results (Webmaster resources, 2014). Usually, phishing webpages are merely accessible for a short period and as a result, many phishing webpages may not be found on the Google index.

29. *Links Pointing to Page:*

The number of links pointing to the webpage indicates its legitimacy level, even if some links are of the same domain. In our datasets and due to its short life span, we find that 98% of phishing dataset items have no links pointing to them. On the other hand, legitimate websites have at least 2 external links pointing to them.

30. *Statistical Report:*

Several parties such as PhishTank and StopBadware formulate numerous statistical reports on phishing websites at every given period of time; some are monthly and others are quarterly. In our research, we used 2 forms of the top ten statistics from PhishTank: "Top 10 Domains" and "Top 10 IPs" according to statistical-reports published in the last three years.

3.2.6 Detecting whether the Entered URL is Phishing or not

The final step of the process is to detecting whether the entered URL is legitimate or phishing by using the dataset obtained in the Dynamic Features extraction step(3.2.5) and using the trained model(3.2.3) which gives the best accuracy.

IV. TESTING AND RESULTS

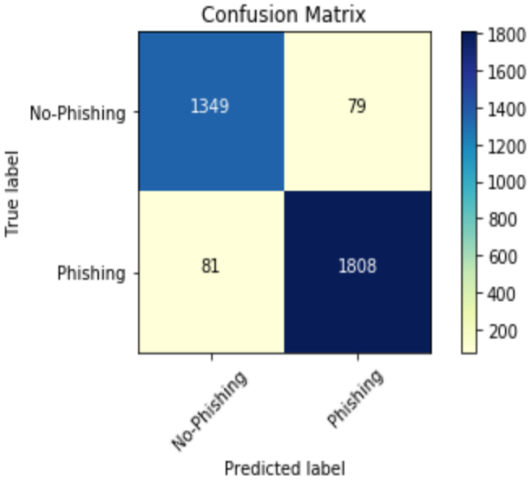

Figure 2: Confusion matrix of ANN

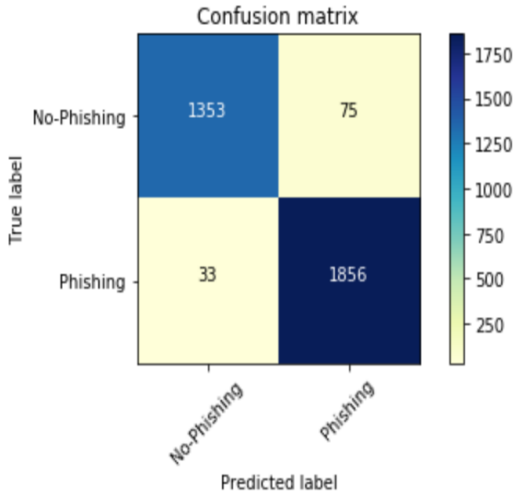

Figure 3: Confusion matrix of Random forest

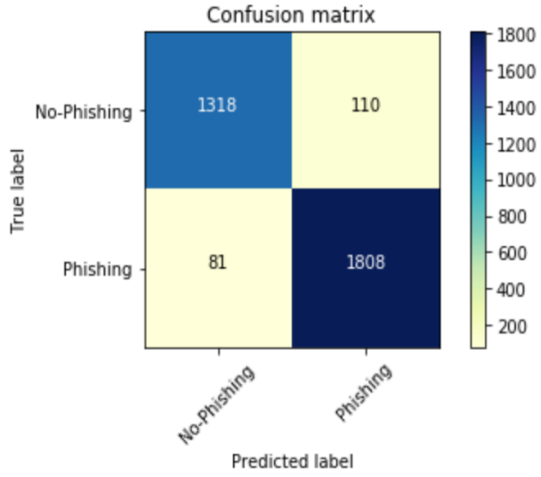

Figure 4: Confusion matrix for Support vecror machine

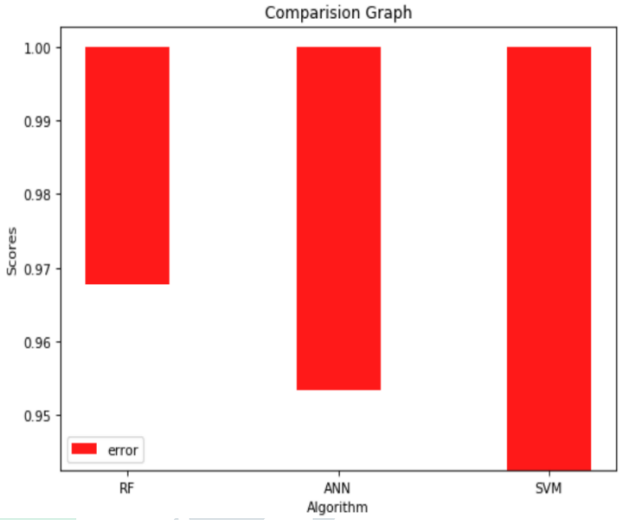

Figure 5: Comparision graph

The figure 2 represents the confusion matrix for artificial neural network with 1349 True Positive, 1808 True Negative, 79 False Positive and 81 False Negative for the testing dataset containing 3317 records.
The figure 3 represents the confusion matrix for random forest with 1353 True Positive, 1856 True Negative, 75 False Positive and 33 False Negative for the testing dataset containing 3317 records.
The figure 4 represents the confusion matrix for support vector machine with 1318 True Positive, 1808 True Negative, 110 False Positive and 81 False Negative for the testing dataset containing 3317 records. And the figure 5 represents the comaprision graph of the three trained models random forest, artificial neural network and support vector machine with accuracy 96.744%, 95.296% and 94.000% respectively.

Table 1: Performance measure of machine learning classifier with wrapper feature selection and without wrapper feature selection

	CLASS	WITH WRAPPER FEATURE SELECTION			WITHOUT FEATURE SELCTION		
		PRECISION (TPR)	RECALL (TNR)	F1-SCORE	PRECISION (TPR)	RECALL (TNR)	F1-SCORE
RANDOM FOREST	0	0.98	0.95	0.96	0.97	0.93	0.95
	1	0.96	0.98	0.97	0.95	0.98	0.96
ARTIFICAL NEURAL NETWORK	0	0.95	0.94	0.95	0.95	0.92	0.94
	1	0.96	0.96	0.96	0.94	0.97	0.96

SUPPORT VECTOR MACHINE	0	0.94	0.92	0.93	0.94	0.92	0.93
	1	0.94	0.92	0.95	0.94	0.96	0.95

Where,

- Precision= True Positive / (True Positive+False Positive)
- Recall= True Positive/ (True Positive+False Negative)
- F1-Score=(2*Recall*Precision) / (Recall+Precision)

- True Positive: Predicted is Positive and it is True
- True Negative: Predicted is Negative and it is True
- False Positive: Predicted is Positive and it is False
- False Negative: Predicted is Negative and it is False

Table 2: Accuracy and error rate

	WITH WRAPPER FEATURE SELECTION			WITHOUT FEATURE SELECTION		
	RANDOM FOREST	NEURAL NETWORK	SUPPORT VECTOR MACHINE	RANDOM FOREST	NEURAL NETWORK	SUPPORT VECTOR MACHINE
ACCURACY %	96.744	95.296	94.000	96.110	94.814	94.332
ERROR %	3.256	4.703	6.000	3.990	5.186	5.668

V. CONCLUSION

The accuracy obtained in the training models of Random forest, Artificial neural network, and Support vector machine are 96.744%, 95.296% and 94% respectively. Therefore the best accuracy obtained is in the Random forest model with wrapper features selection, which is used in the final detection of phishing websites.

The most important way to protect the user from phishing attack is the education awareness. Internet users must be aware of all security tips which are given by experts. Every user should also be trained not to blindly follow the links to websites where they have to enter their sensitive information. It is essential to check the URL before entering the website.

VI. FUTURE ENHANCEMENTS

In future system can upgrade to detect phishing web pages automatically by running in the background for user sessions so that the system can prevent the phishing attacks very efficiently.

REFERENCES

[1] Routhu Srinivasa Rao and Syed Taqi Ali, 2015 "PhishShield: A Desktop Application to Detect Phishing Webpages through Heuristic Approach" Department of Computer Engineering, National Institute of Technology, Kurukshetra , Haryana, India, Procedia Computer Science 54,(147 – 156).
[2] Hemali Sampat, Manisha Saharkar, Ajay Pandey, Hezal Lopes, 2018" Detection of Phishing Website Using Machine Learning" Department of Computer Engineering, Universal College of Engineering, Vasai , Maharashtra, India,IRJET Volume-5 Issue-3.
[3] S. Carolin Jeeva and Elijah Blessing Raj singh. 2016 "Intelligent phishing url detection using association rule mining" Department of Computer Applications ,Karunya University, Coimbatore, India.
[4] Sa'id Abdullah Al-Saaidah, 2017 "Detecting Phishing Emails Using Machine Learning Techniques" Department of Computer Science Faculty of Information Technology Middle East University.
[5] Ram B. Basnet1, Andrew H. Sung, Quingzhong Liu, 2014 "Learning to Detect Phishing URLs" Colorado Mesa University, 1100 North Ave. Grand Jct. CO 81501, USA IJRET.
[6] Ankith Kumar jain and BB Guptha,2017 "Phishing Detection: Analysis of Visual Similarity Based Approaches" National Institute of Technology, Kurukshetra, India, Hindawi Security and Communication NetworksVolume 2017.
[7] Jian Mao, Jingdong Bian, Wenqian Tian, Shishi Zhu, Tao Wei, Aili Li and Zhenkai Liang, 2018 "Detecting Phishing Websites via Aggregation Analysis of Page Layouts" Procedia Computer Science 129 (2018) 224–230.
[8] R. Kiruthiga, D. Akila,2019 "Phishing Websites Detection Using Machine Learning" (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S11.
[9] Moitrayee Chatterjee and Akbar Siami Namin, 2019**"**Detecting Phishing Websites through Deep Reinforcement Learning"IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC).
[10] M E Pratiwi ,2018 "Phishing Site Detection Analysis Using Artificial Neural Network" Journal of Physics: Conference Series-1140 (2018) doi:10.1088/1742-6596/1140/1/012048.