RESEARCH ARTICLE OPEN ACCESS

Phishing Website Detection using Machine Learning Algorithms

Mrs B Vijaya MCA.,M.Tech.,[1] , Abboori Sekhar [2]
[1] Associate Professor, Department of Computer Applications
[2] Student, Department of Computer Applications
[1],[2] Chadalawada Ramanamma Engineering College

ABSTRACT
Phishing attacks are one of the trendy cyber attacks that include socially formed messages communicated to experts who try to fool consumers to report their sensitivity data, users' emails are the most common contact medium for such messages. This is what you are talking about. Paper proposes a smart classification model for the identification by knowledge discovery of phishing email techniques for data mining and text handling. This paper presents the idea of weighting phishing words that assesses in each email the weight of the phishing words. The preparatory step is strengthened by to enrich the model with word synonymes, the implementation of text stemming and WordNet ontology. The model was used the exploration of information using five popular algorithms and has achieved remarkable results classification exactness improvements; the Random Forest algorithm has been used to reach 99.1% precision and 98.4% use J48, which is the best in our experience accuracy rate for an accredited data set.
Keywords: - cyber-attacks sensitivity data, users' emails are the most common contact medium for such messages.

I. INTRODUCTION
Phishing is a wrongdoing in which a culprit sends the phony email, which seems to come from famous and confided in brand or association, requesting to include individual qualification like bank secret phrase, username, telephone number, address, Mastercard subtleties, etc [1–4]. The phony messages frequently look incredibly authentic, and even the site where the Internet client is approached to include individual data likewise appears to be like genuine one. Phishing messages engender over email, SMS, moment couriers, informal communication destinations, VoIP, etc, however email is the famous method to play out this assault and 65% of the absolute phishing assault is accomplished by visiting the hyperlink appended to the email [5]. Also, skewer phishing assault is turning out to be well known these days. Business email bargain (BEC) is seen as a significant Internet danger in 2015 [6]. In BEC, the interloper utilizes skewer phishing strategies to trick associations and Internet people. More complex lance phishing assaults [7–9] focused on specific individual or gatherings inside the association. Phishing is allegorically like fishing in the water, yet as opposed to attempting to get a fish, assailants attempt to take purchaser's very own data [10, 11]. At the point when a client opens a phony site page and enters the username and ensured secret word, the qualifications of the client are procured by the aggressor which can be utilized for pernicious purposes [12–22]. Phishing sites look fundamentally the same as in appearance to their comparing real sites to pull in enormous number of Internet clients. Late improvements in phishing recognition have prompted the development of various new visual closeness based methodologies. Visual likeness based methodologies look at the visual appearance of the dubious site to its relating genuine site by utilizing different boundaries. The new advances in web and portable innovation pulled in most business organizations to offer their administrations web based, including banks, stocks and internet business suppliers. As individuals progressively depend on Internet administrations to do their exchanges, Internet extortion turns into an incredible danger to individuals' security and wellbeing. Phishing is one of the principle kinds of Internet misrepresentation; which depends on tricking clients to share or proclaim their private data (counting passwords and Mastercard numbers), phishing could be characterized as a digital assault that imparts socially-designed messages to people through electronic correspondence channels (email, SMS, call) in request to convince them to do certain activities (enter accreditations, charge card number, …) for the aggressors advantage; such activities could be convincing an internet business site client to enter his certifications to a phony site (oversaw by the assailant) like the first site and afterward the assailant utilizes them to mimic the client. To convince the casualty client to login to a particularly phony site, the socially designed message attracts a hallucination to the client that he needs to perform such activity, for example, notice the client about record suspension or that the site administrator is mentioning him to reset his secret phrase [1]. Phishing assaults utilize email messages and sites that

are planned in an expert way to be like messages and sites from real foundations and associations (normally the client is a client for those associations), to convince clients into revealing their own or monetary data. The aggressor would then be able to utilize gathered touchy client data for his advantage. Clients can be fooled into revealing their data either by giving touchy data through a web structure, answering to ridiculed messages, or downloading and introducing Trojans, which search clients' PCs or screen clients' online exercises to get data. Phishing assaults have consistently expanded to coordinate the development of electronic trade, as of late taking on pandemic extents; the Anti Phishing Work Group (APWG) report of 2015 [2] announced that the complete number of extraordinary phishing destinations identified from Quarter1 through Quarter3 of 2015 was 630,494, while The quantity of special phishing reports submitted to APWG from quarter 1 through quarter3 was 1,033,698. As indicated by a new report from Google [3], 45% of phishing sites tricked their objective casualties into pronouncing their passwords, and got their secret phrase changed by the aggressor inside 30 minutes after their records were captured. The aggressors likewise abused the casualties' records in tricking others in the casualty's contact list through speaking with them utilizing the commandeered accounts; the examination presumed that those individuals are multiple times bound to be captured when the assailants utilized the casualty's record to speak with them, and this is a normal outcome as the correspondence is gotten from a confided in record. Numerous scientists have examined the phishing issue and proposed an assortment of answers for battle phishing assaults. The main class of proposed arrangements deals with the standard of distinguishing phishing assaults and cautioning the client or keeping him from making moves that could bring about trading off his private information, most recent exploration recommendations in this classification incorporate [4] [5] [6] [7] [8] [9]. The second class of proposed phishing arrangements depend on making sure about the login measure by adding a subsequent verification factor with the end goal that taking the client's accreditations isn't enough for an assailant to bargain the casualty's record except if he additionally has the second validation factor, those recommendations incorporate [10] [11] [12] [13] [14] [15] [16] [17] [18] [19]. Our concentration in this paper is to assemble a keen classifier at the email level that is able to do recognizing phishing messages as a beginning phase in the phishing fighting cycle; we accept that recognizing phishing messages can make the web clients safer by dispensing with those messages furthermore, not depending on the clients' watchfulness to shield them from phishing assaults; numerous investigations inferred that relying upon human components is definitely not a favored choice for fighting phishing assaults; particularly for cutting edge and decidedly ready phishing assaults that are ceaselessly adjusting themselves to known safeguard systems [20] [21].

II. RELATED WORK

First phishing assault was seen on America online organization frameworks (AOL) in the mid 1990s [20] where numerous false clients enlisted on AOL site with counterfeit Mastercard subtleties. AOL passed these phony records with a basic legitimacy test without confirming the authenticity of the Mastercard. After enactment of the phony record, aggressors got to the assets of America online framework. At the hour of charging, AOL verified that the records were false, and related Mastercards were likewise not substantial; in this way AOL stopped these records right away. After this episode, AOL took measures to forestall this sort of assault by confirming the genuineness of Mastercard and related charging character, which likewise empowered the assailants to change their method of acquiring AOL accounts. Rather than making a phony record, assailants would take the individual data of enrolled AOL client. Aggressors reached enlisted AOL clients through moment courier or email and approached them to confirm the secret word for security purposes. Email and texts seemed to come from an AOL worker. Numerous clients gave their passwords and other individual data to the aggressors. The aggressors at that point utilized the differently charged segments of America online site in the interest of a genuine client. Besides, an assailant no longer confines themselves to disguising America online site however effectively disguise an enormous number of monetary and electronic trade sites.

Another examination [22] applied a two-stage grouping model of messages; in the principal stage a bunch of arrangement calculations (C5.0, Naive Bayes, SVM, Linear Regression and K-Nearest Neighbors) are utilized to arrange genuine and phishing messages, normal assessment measurements are used to assess every calculation including exactness, accuracy, review and F-score, the calculation with best order results was C5.0 with a normal exactness pace of 97.15%, normal exactness of 98.56%, normal review of 95.64% and normal F-score of 97.08%. in the second stage, the messages that were named real in the main stage were contribution to an outfit classifier.

The authors in [13] proposed an email order model that misuses 23 catchphrases extricated from the email body, the proposed model was tried utilizing a bunch of arrangement calculations, counting multilayer perceptron, choice trees, uphold area machine, probabilistic neural net, hereditary programming, and calculated relapse. The best order result was accomplished utilizing
hereditary programming with an arrangement precision of 98.12%.

The investigation [14] applies the Bayesian classifier for phishing email recognition, assessed regarding exactness, mistake, time, accuracy and review. The model brought about precision of 96.46%.

The authors in [15] applied Support Vector Machine classifier to arrange messages utilizing a bunch of structure-based and conduct based highlights. The model accomplished 97.25% precision in outcomes, anyway its shortcoming is in its generally little preparing dataset (1000 messages with half spam and half ham).

The authors in [16] proposed an email characterization calculation by coordinating Bayesian Classifier furthermore, phishing URLs location utilizing Decision Tree C4.5, their methodology accomplished 95.54 % exactness, which is superior to the precision of 94.86% that was accomplished utilizing Bayesian classifier.

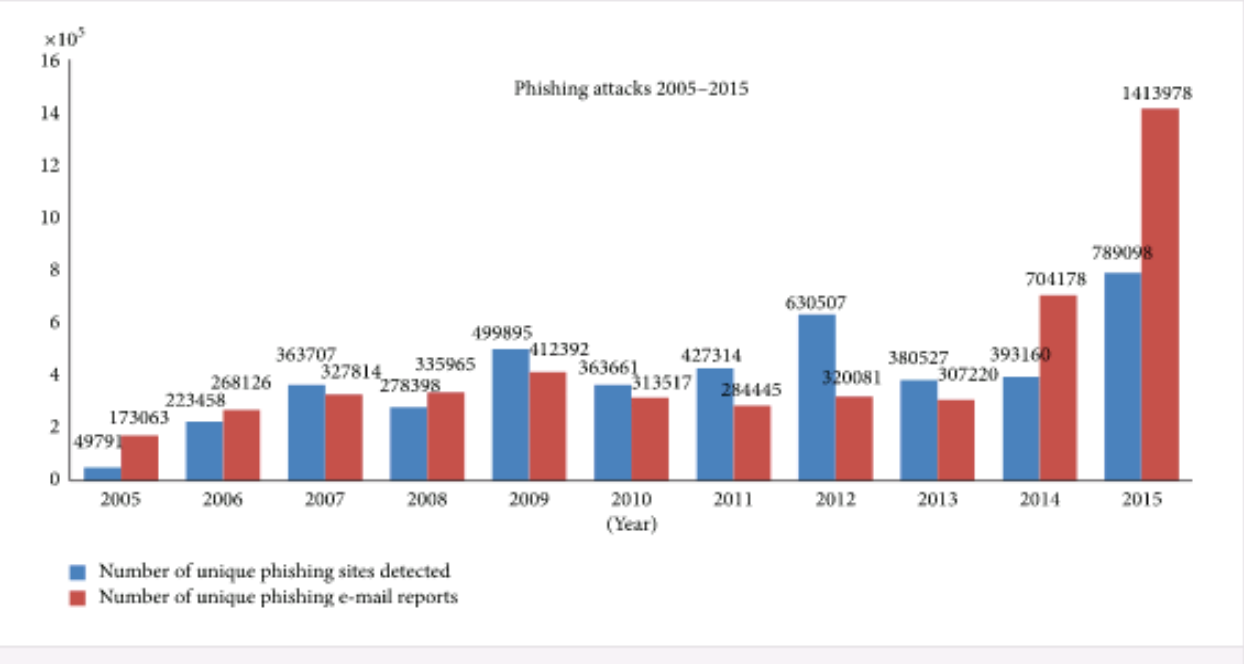
The examination in [17] utilized Random Forest and Partial Decision Tree calculation for spam email grouping, the creators applied a bunch of highlight determination techniques in the pre-preparing step counting Chi-square and Information pick up, they accomplished exactness of 96.181% with Random Woodland and 95.093% with Part.

The authors in [18] proposed a program information based compound methodology for recognizing phishing assaults, the proposed model investigations web URLs utilizing parsing and uses a bunch of kept up information bases which store the recently visited URLs and recently identified phishing URLs. The test results demonstrated 96.94% exactness in distinguishing phishing URLs with a little trade off in corrupting the program speed.

III. STATISTICS

As indicated by Internet world details [38], all out quantities of Internet clients overall are 2.97 billion of every 2014; that is, over 38% of the total populace utilizes Internet. Programmers exploit the unreliable Internet framework and can trick ignorant clients to succumb to phishing tricks. Phishing email is utilized to dupe the two people and monetary associations on the Internet. The Anti-Phishing Working Group (APWG) [19] is a global consortium which is committed to advancing exploration, training, and law authorization to dispose of online extortion and digital wrongdoing.

In 2012, all out phishing assault expanded by 160% more than 2011, meaning a record year in phishing volumes. The complete phishing assaults recognized in 2013 were around 450000 and prompted monetary misfortunes more than 5.9 billion dollars [19]. All out assault increments by 1% in 2013 when contrasted with 2012. The absolute number of phishing assaults saw in Q1 (first quarter) of 2014 was 125,215, a 10.7 percent expansion over Q4 (final quarter) of 2013. Over 55% of phishing websites contain the name of the objective site in some structure to trick clients and 99.4% of phishing websites utilize port 80 [20]. As per the APWG report in the primary quarter of 2014, second most elevated number of phishing assaults ever recorded was among January and March 2014 [20] and installment administrations are the most focused on industry. During the second 50% of 2014, 123,972 exceptional phishing assaults were noticed [21]. In the year 2011, complete monetary misfortunes were 1.2 billion, and they rose to 5.9 billion dollars in 2013. The monetary misfortunes due to phishing assault in 2014 and 2015 were 4.5 and 4.6, individually, as appeared in Figure 1 [22]. The development of phishing assaults from 2005 to 2015 is appeared in Figure



IV. PHISHING MECHANISM

The phishing mechanism is appeared in Figure. The phony website is the clone of focused certified website, and it generally contains some information fields (e.g., text box). At the point when the client presents his/her own subtleties, the data is moved to the aggressor. An aggressor takes the accreditation of the honest client by performing following advances:

Construction of Phishing Site

In the initial step aggressor recognizes the objective as a notable association. Subsequently, aggressor gathers the point by point data about the association by visiting their website. The assailant then uses this data to develop the phony website. URL Sending. In this progression, assailant makes a counterfeit email and sends it to the large number of clients. Assailant joined the URL of the phony website in the counterfeit email. On account of lance phishing assault, an assailant sends the email to chosen clients. An assailant can likewise spread the connection of phishing website with the assistance of web journals, discussion, etc

Stealing of the Credentials

At the point when client taps on appended URL, therefore, counterfeit site is opened in the internet browser. The phony website contains a phony login structure which is utilized to take the accreditation of an honest client. Furthermore, assailant can get to the data filled by the client.

Identity Theft

Assailant utilizes this qualification of malignant purposes. For instance, assailant buys something by utilizing charge card subtleties of the client.

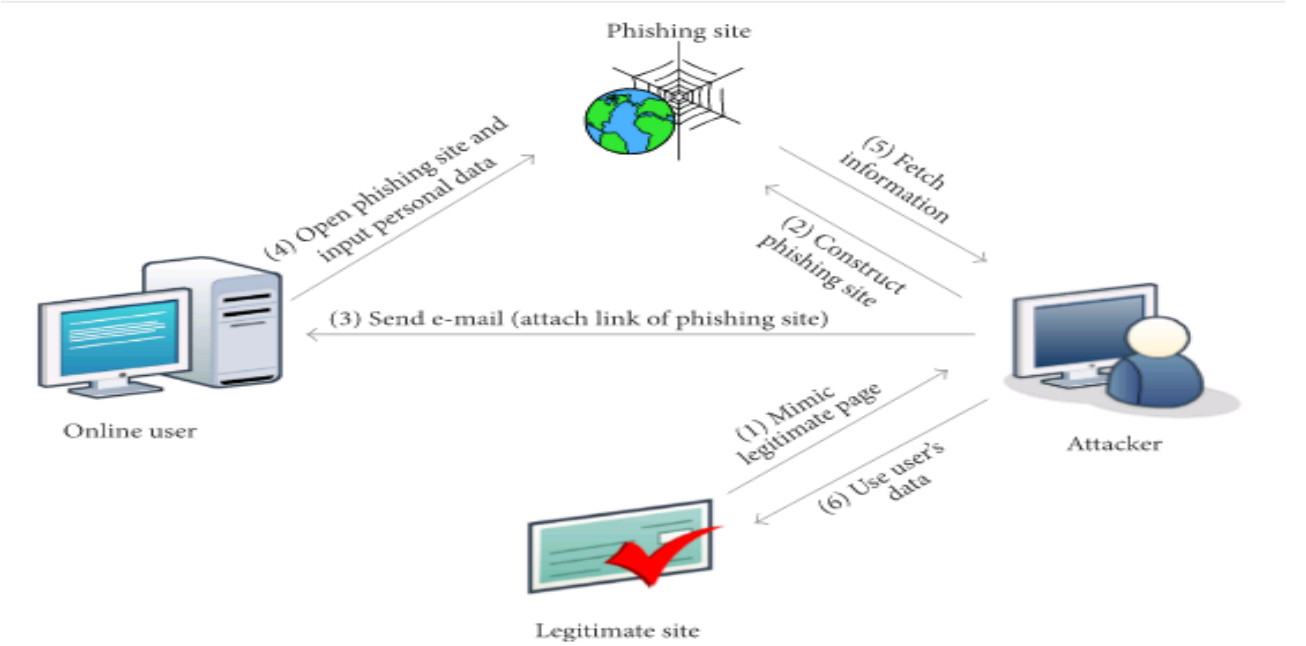
Support Vector Machine Algorithm

"Support Vector Machine" (SVM) is a supervised algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well

Training Model for SVM

Input: D=[X,Y]; X(array of input with m features), Y(array of class labels)
Y=array(C) // Class label
Output: Find the performance of the system
function train_svm(X,Y, number_of_runs)
initialize:learning_rate=Math.random();
for learning_ rate **in**number_of_runs
error=0;
for i **in** X
if (Y[i] *(X[i]*w))<1 **then**
update : w=w + learning_rate * ((X[i]*Y[i])*(-2*(1/number_of_runs)*w)
else
update: w=w+learing_rate *(-2*(1/number_of_runs)*w)
end if

end
end

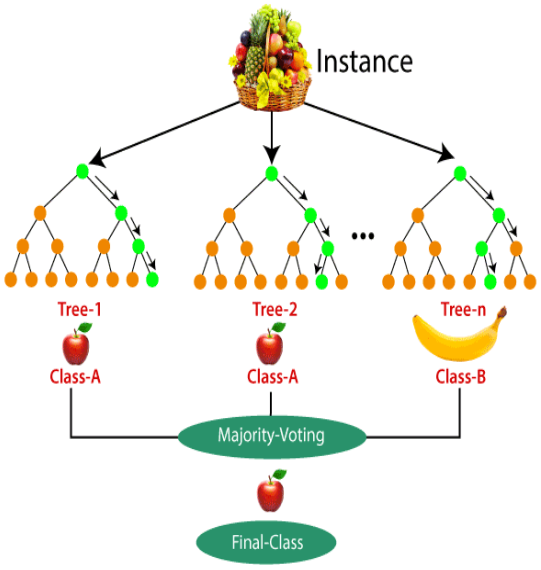


Random Forest Algorithm

Random Forest algorithm is a supervised classification algorithm. We can see it from its name, which is to create a forest by some way and make it random. There is a direct relationship between the number of trees in the forest and the results it can get: the larger the number of trees, the more accurate the result. But one thing to note is that creating the forest is not the same as constructing the decision with information gain or gain index approach.

There are two stages in Random Forest algorithm, one is random forest creation, the other is to make a prediction from the random forest classifier created in the first stage. Here shows the Random Forest creation pseudocode:

1. Randomly select "**K**" features from total "**m**" features where **k << m**
2. Among the "**K**" features, calculate the node "**d**" using the best split point
3. Split the node into **daughter nodes** using the **best split**
4. Repeat the **a to c** steps until "l" number of nodes has been reached
5. Build forest by repeating steps **a to d** for "n" number times to create **"n" number of trees**



Require: Initially the tree has exactly one leaf (TreeRoot) which covers the whole space
Require: The dimensionality of the input, D. Parameters λ, m and τ.
SelectCandidateSplitDimensions(TreeRoot, min(1 + Poisson(λ), D))
for t = 1 . . . do

```
    Receive (Xt, Yt, It) from the environment At ←
    leaf containing Xt
    if It = estimation then
    UpdateEstimationStatistics(At, (Xt, Yt)) for all S
    ∈ CandidateSplits(At) do
    for all A ∈ CandidateChildren(S) do
    if Xt ∈ A then UpdateEstimationStatistics(A, (Xt,
    Yt))   end if
    end for
    end for
else if It = structure then
    if At has fewer than m candidate split points then
    for all d ∈ CandidateSplitDimensions(At) do
    CreateCandidateSplit(At, d, πdXt)
    end for
    end if
    for all S ∈ CandidateSplits(At) do
    for all A ∈ CandidateChildren(S) do
     if Xt ∈ A then UpdateStructuralStatistics(A, (Xt,
      Yt))      end if
     end for
    end for
     if CanSplit(At) then
      if ShouldSplit(At) then
           Split(At)
      else if MustSplit(At) then
       Split(At)
      end if
     end if
    end if
    end for
```

5. Proposed Model

The proposed approach for phishing email grouping utilizes the model of Knowledge Discovery (KD) and data digging for building a shrewd email classifier that can group another email message as a real or spam; the proposed model is worked by applying the iterative strides of KD to distinguish and separate helpful features from a preparation email data set, the features are then taken care of to a gathering of data mining calculations to recognize the best classifier. The proposed model for email arrangement uses etymological processing methods and ontologies to upgrade the closeness between messages with comparable semantic term meaning additionally the standard of term report recurrence is applied in weighting the phishing terms in each email with the end goal that messages phishing terms weighting helps in separating phishing from real messages. The proposed model additionally decreased the quantity of features utilized in the order cycle into 16 features in particular; which improves the characterization execution and proficiency and limits the commotion of including numerous features and thus improves the arrangement precision. These upgrades and are examined in detail in the accompanying subsections.

Knowledge Discovery Model

Knowledge discovery is the way toward removing or finding designs from data, the separated examples should be novel, substantial, valuable and understandable [9]. The KD cycle is completed utilizing a bunch of iterative strides as portrayed in figure. The means are started by understanding the issue and the data, trailed by a data pre-processing stage to prepare it for the data mining venture through which the objective knowledge is found, assessed and then presented as a helpful and simple to utilize data.

Data Collection

The initial phase in building the proposed phishing email classifier is picking the appropriate preparing data set which is a genuine example of existing messages that comprises of both phishing and real messages (otherwise called spam and ham messages). The preparation data set will be utilized to find conceivably predictive connections that will fill in as building blocks in the classifier. Our preparation data set comprises of 10538 messages including 5940 ham messages from spam professional killer project [3] and 4598 spam messages from Nazario phishing corpus [3].

Data Pre-processing and features extraction

In this progression the messages in the preparation data set are prepared and sifted with the end goal that they can be changed into a data design that is effectively and successfully prepared in ensuing strides of building the classifier. The messages in our picked preparing data set are accessible in plain content organization which should be pre-handled and changed into EML design (Microsoft Outlook Express document augmentation) that is interoperable with the java mail bundle that will be utilized to extricate the email features.

6. Result and Discussion

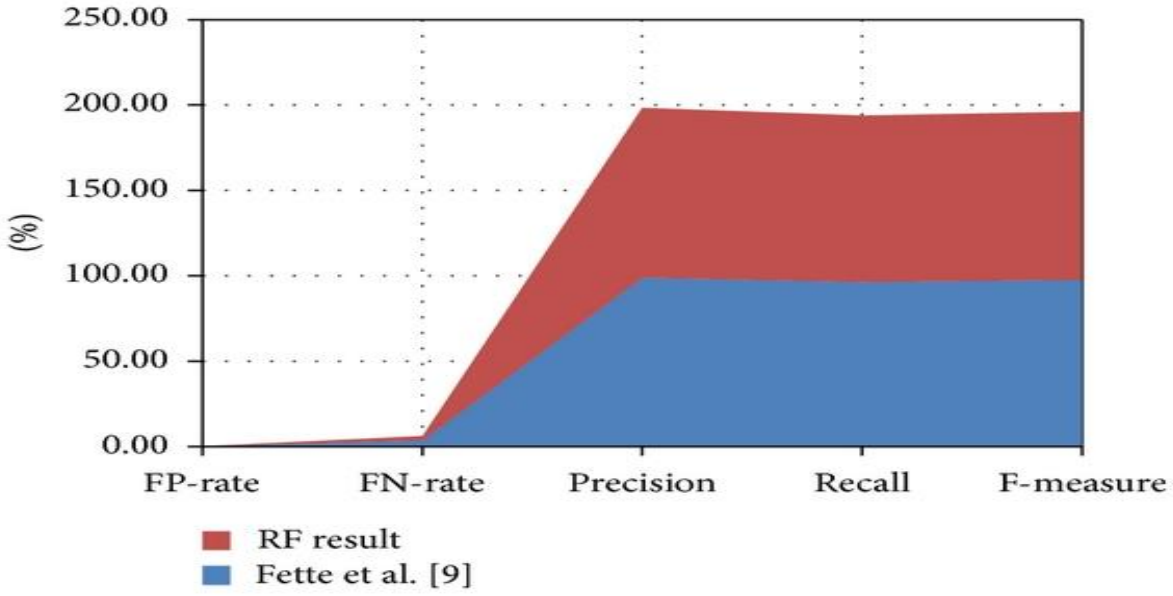
Machine learning involves two major phases: the training phase and the testing phase. The predictive accuracy of the classifier solely depends on the information gained during the training process; if the information gained (IG) is low, the predictive accuracy is going to be low, but if the IG is high, then the classifier's accuracy will also be high.

As stated above, we used 10-fold cross validation. In our random forest classification, before the decision trees are constructed, the information gained for all the 15 features is calculated (using the IG method explained by Mitchell [24]) and the features with the best eight IG are selected and used for constructing the decision trees; the mode vote (from all the trees)

is then calculated and used for the email prediction. Information gain is one of the feature ranking metric highly used in many text classification problems today. More details about our algorithm are described in the next section below.

We tested our method using varied dataset sizes, this was done to know the performance of the algorithm on both small and large datasets. As shown in the table, the algorithm performed best when tested on the dataset that has the largest size (having an overall accuracy of 99.7%, FN rate of 2.50%, and FP rate of 0.06%); this implies that our method will work effectively if applied to real world dataset, which is usually large in size. Our method also achieved a higher prediction accuracy (99.7%) compared to an accuracy of 97% achieved by Fette et al. [9].

example, text comparability, textual style tone, text dimension, and pictures present in the site page. Text based comparability approaches are moderately quick, yet they can't identify phishing assault if the content is supplanted with some picture. Picture handling based methodologies have high exactness rate while they are unpredictable in nature and are tedious. Moreover, the greater part of the work is done disconnected. These include information assortment and profile-creation stages to be finished first. A near table is ready for simple looking at the preferences and disadvantages of the accessible methodologies. No single strategy is sufficient for receiving it for phishing identification purposes. Identification of phishing sites with high exactness is as yet an open test for additional innovative work.



V. CONCLUSION

Phishing is a horrifying danger in the web security space. In this assault, the client inputs his/her own data to a phony site which resembles a real one. We have introduced an overview on phishing location approaches dependent on visual similitude. This review gives a superior comprehension of phishing site, different arrangement, and future degree in phishing recognition. Numerous methodologies are talked about in this paper for phishing discovery; anyway the vast majority of the methodologies actually have constraints like precision, the countermeasure against new phishing sites, neglecting to distinguish inserted objects, etc. These methodologies utilize different highlights of a website page to recognize phishing assaults, for

Technique	FP-Rate	FN-Rate	Precision	Recall	F-Measure
Fette et al. [9]	0.13%	3.62%	98.92%	96.38%	97.64%
RF Result	0.06%	2.50%	99.47%	97.50%	98.45%

VI. REFERENCES

1. Rao, Y. Narasimha, et al. "Mimicked Web Page Detection over Internet." International Journal of Electronics Communication and Computer Engineering 5.1 (2014): 104.
2. Rao, Y. Narasimha, and Bhavya Nalamothu. "ENHANCED STEGANOGRAPHIC SCHEMES TO SECURE DATA IN IOT." *Journal of Natural Remedies* 21.4 (2020): 88-96.
3. A. K. Jain and B. B. Gupta, "Comparative analysis of features based machine learning approaches for phishing detection," in Proceedings of the 10th INDIA-COM, New Delhi, India, 2016.
4. G. Weaver, A. Furr, and R. Norton, Deception of Phishing: Studying the Techniques of Social Engineering by Analyzing Modern-Day Phishing Attacks on Universities, 2016.
5. Kaspersky Lab, "Spam in January 2012 love, politics and sport," 2013, http://www.kaspersky.com/about/news/spam/2012/Spam_in_January_2012_Love_Politics_and_Sport.
6. APWG Q1-Q3 Report, 2015, http://docs.apwg.org/reports/apwg_trends_report_q1-q3_2015.pdf.
7. B. Parmar, "Protecting against spear-phishing," Computer Fraud & Security, vol. 2012, no. 1, pp. 8–11, 2012.
8. W. Jingguo, T. Herath, C. Rui, A. Vishwanath, and H. R. Rao, "Phishing susceptibility: an investigation into the processing of a targeted spear phishing e-mail," IEEE Transactions on Professional Communication, vol. 55, no. 4, pp. 345–362, 2012.
9. T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," Communications of the ACM, vol. 50, no. 10, pp. 94–100, 2007.
10. C. H. Hsu, P. Wang, and S. Pu, "Identify fixed-path phishing attack by STC," in Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS '11), pp. 172–175, ACM, Perth, Australia, September 2011.
11. N. A. G. Arachchilage and M. Cole, "Designing a mobile game for home computer users to protect against phishing attacks," https://arxiv.org/abs/1602.03929.
12. R. Dhamija and J. D. Tygar, "The battle against phishing: Dynamic security skins," in Proceedings of the Symposium on Usable Privacy and Security (SOUPS '05), pp. 77–88, July 2005.
13. S. Sheng, B. Magnien, P. Kumaraguru et al., "Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish," in Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07), Pittsburgh, Pa, USA, July 2007.
14. K.-P. Yee and K. Sitaker, "Passpet: convenient password management and phishing protection," in Proceedings of the 2nd Symposium on Usable Privacy and Security (SOUPS '06), pp. 32–43, ACM, Pittsburgh, Pa, USA, July 2006.
15. E. Gal´an and J.C. Hern andez Castro and A. Alcaide and A. Ribagorda, "A Strong Authentication Protocol based on Portable One–Time Dynamic URLs", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2010.
16. Mengjun Xie, Yanyan Li, Kenji Yoshigoe, Remzi Seker, Jiang Bian, "CamAuth: Securing Web Authentication with Camera", IEEE 16th International Symposium on High Assurance Systems Engineering, 2015.
17. http://www.google.com/landing/2step/. Accessed June 2016.
18. A. Czeskis, M. Dietz, T. Kohno, D. Wallach, and D. Balfanz, "Strengthening

user authentication through opportunistic cryptographic identity assertions," in Proceedings of the 2012 ACM conference on Computer and communications security, ser. CCS '12, 2012, pp. 404–414.

19. Ben Dodson, Debangsu Sengupta, Dan Boneh, and Monica S. Lam, "Secure, Consumer-Friendly Web Authentication and Payments with a Phone". http://mobisocial.stanford.edu/papers/mobicase10s.pdf, Accessed June 2016.
20. Rachna Dhamija, J. D. Tygar, Marti Hearst, "Why Phishing Works", CHI-2006: Conference on Human Factors in Computing Systems, 2006.
21. Julie S. Downs, Mandy B. Holbrook, Lorrie Faith Cranor, "Decision Strategies and Susceptibility to Phishing", Symposium On Usable Privacy and Security (SOUPS), July 12-14, 2006, Pittsburgh, PA, USA
22. F. Toolan and J. Carthy, "Phishing detection using classifier ensembles," in eCrime Researchers Summit, 2009. eCRIME'09. IEEE, 2009, pp.1–9.
23. Mayank Pandey and Vadlamani Ravi, "Detecting phishing e-mails using Text and Data mining", IEEE International Conference on Computational Intelligence and Computing Research 2012.
24. Sunil B. Rathod, Tareek M. Pattewar, "Content Based Spam Detection in Email using Bayesian Classifier", IEEE ICCSP conference, 2015.
25. Lew May Form, Kang Leng Chiew, San Nah Szeand Wei King Tiong, "Phishing Email Detection Technique by using Hybrid Features", IT in Asia (CITA), 9th International Conference, 2015.
26. Tareek M. Pattewar, Sunil B. Rathod, "A Comparative Performance Evaluation of Content Based Spam and Malicious URL Detection in E-mail", IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), 2015.
27. Prajakta Ozarkar, & Dr. Manasi Patwardhan," Efficient Spam Classification by Appropriate Feature Selection", International Journal of Computer Engineering and Technology (IJCET), ISSN 0976 – 6375(Online) Volume 4, Issue 3, May – June (2013).
28. Gaurav Kumar Tak1 and Gaurav Ojha2, "MULTI-LEVEL PARSING BASED APPROACH AGAINST PHISHING ATTACKS WITH THE HELP OF KNOWLEDGE BASES", International Journal of Network Security & Its Applications (IJNSA), Vol.5, No.6, November 2013
29. Usama Fayyad, Gregory Piatetsky Shapiro and Padhraic Smyth "Knowledge Discovery and Data Mining: Towards a Unifying Framework", KDD-96 Proceedings, 1996.