




Article

An Effective and Secure Mechanism for Phishing Attacks Using a Machine Learning Approach

Gori Mohamed ¹, J. Visumathi ², Miroslav Mahdal ^{3,*} , Jose Anand ⁴  and Muniyandy Elangovan ^{5,*} 

¹ Department of Information and Communication Engineering, Anna University, Chennai 600 025, India; drghorimohamed@gmail.com

² Department of Computer Science and Engineering, Veltech Rangarajan Dr Sangunthala R&D Institute of Science and Technology, Chennai 600 092, India; jsvisu@gmail.com

³ Department of Control Systems and Instrumentation, Faculty of Mechanical Engineering, VSB-Technical University of Ostrava, 17. Listopadu 2172/15, 708 00 Ostrava, Czech Republic

⁴ Department of Electronics and Communication Engineering, KCG College of Technology Karapakkam, Chennai 600 097, India; joseanandme@yahoo.co.in

⁵ Department of R&D, Bond Marine Consultancy, London EC1V 2NX, UK

* Correspondence: miroslav.mahdal@vsb.cz (M.M.); muniyandy.e@gmail.com (M.E.)

Abstract: Phishing is one of the biggest crimes in the world and involves the theft of the user's sensitive data. Usually, phishing websites target individuals' websites, organizations, sites for cloud storage, and government websites. Most users, while surfing the internet, are unaware of phishing attacks. Many existing phishing approaches have failed in providing a useful way to the issues facing e-mails attacks. Currently, hardware-based phishing approaches are used to face software attacks. Due to the rise in these kinds of problems, the proposed work focused on a three-stage phishing series attack for precisely detecting the problems in a content-based manner as a phishing attack mechanism. There were three input values—uniform resource locators and traffic and web content based on features of a phishing attack and non-attack of phishing website technique features. To implement the proposed phishing attack mechanism, a dataset is collected from recent phishing cases. It was found that real phishing cases give a higher accuracy on both zero-day phishing attacks and in phishing attack detection. Three different classifiers were used to determine classification accuracy in detecting phishing, resulting in a classification accuracy of 95.18%, 85.45%, and 78.89%, for NN, SVM, and RF, respectively. The results suggest that a machine learning approach is best for detecting phishing.

Keywords: phishing; attack detection; web crawler; heuristic analysis; machine learning classification



Citation: Mohamed, G.; Visumathi, J.; Mahdal, M.; Anand, J.; Elangovan, M. An Effective and Secure Mechanism for Phishing Attacks Using a Machine Learning Approach.

Processes **2022**, *10*, 1356. <https://doi.org/10.3390/pr10071356>

Received: 28 June 2022

Accepted: 6 July 2022

Published: 12 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the modern era of networks, almost every industry uses the internet. Many different security attacks affect businesses. Among the main attacks is phishing. Phishing threats are conducted by e-mail spoofing and similar webpage functioning. The phisher can perform attacks with the help of spoofed e-mails and by copying website design. Based on the internet, the phisher can hack the user's personal belongings. Phishing is an offence implemented by technical and social technology and hacking user identity information and banking information. Phishing threats are enabled by user weakness and the development of sophisticated mechanisms by phishers [1].

When working on the internet, users need to enter user data such as personal information and banking information. These attacks are used to steal the user's data. Phishing attacks are increasing. Phishing websites look the same as the original websites [2]. A group named Anti-Phishing was formed to control phishing attacks. A report of that group says that phishing activities are increasing. The main target of phishers is to attack the victim's e-mails, messages and phone calls. There are many kinds of phishing, such as deceptive

phishing in which the attacker's focus is on the organization in which the employees work. Deceptive phishing is easily implemented by using the URL to distinguish genuine links from the scammer. Phishing Spear is a kind of phishing in which attacks are conducted through e-mail by targeting and collecting data about the entity on Facebook. Attackers are targeted through emails by crafting a positioning attack on DNS. User IP address is easily identified by attackers through DNS and users' IP addresses are easily redirected to malicious websites. A new type of phishing was discovered which is conducted through dropbox—attackers want to steal the dropbox files from users. Phishing attackers create a fake dropbox signature and then this is passed to the dropbox of users. Phishers can easily steal files and users' credentials which are hosted on the website. Google docs phishing was derived from dropbox phishing, through which attackers can easily target victims. Attackers initiate on SaaS or webmail by stealing sensitive data as their primary goal. To integrate e-mail addresses into the system, an anti-phishing simulator was designed to prevent serious threats by catching malicious emails arriving in the system. This system also helps to evaluate keywords in the existing database and to determine the contents of the database.

1.1. Existing Issues of Phishing on Website

In 2011, Prevost et al. identified phishing based on more than 25 features on a heuristic webpage. The main disadvantage of this work is that when there are changes in webpages, this method does not work, i.e., it is not robust. It also requires more heuristic parameters [2]. In 2016, Moghimi et al. developed rule-based phishing techniques, based on the parameters of a webpage and applied the rules for retrieving the hidden information in the webpage. The major drawback of this algorithm is that it is not reliable for identifying the phisher [3]. In 2010, Prakash et al. applied predictive blacklist techniques to identify and delete the correct blacklist using the PhishNet database. This algorithm fails to detect the 0th-day phishing [4].

1.2. Existing Issues of Phishing in E-Mail

In 2006, Lyon et al. proposed the authentication for domain level by securely sending the data by e-mail. The main drawback of this algorithm is that it works only when both the sender and receiver have the same device [5]. Chen et al. developed the LinkGuard algorithm. This algorithm is capable of detecting the actual domain of phishers and also detecting the phisher link URL. The main drawback of this algorithm is that it is applicable only for e-mail and not for detecting phishers on the web [6]. In 2009, Gansterer et al. applied a machine learning-based K-nearest learning approach to detect phishing based on the parameters of e-mail and ranked those parameters. The main limitations are that it attained a high false-positive rate for spam detecting filters [7,8].

A phishing attack is a dangerous threat. For example, at Delaware University, nearly 75 thousand people faced this problem as phishers stole the personal information of teachers, students, researchers and faculty through websites [9]. In 2001, the first online gold fraud was investigated. This was similar to a phisher sending spam mail to increase their network. Since the number of phishing attacks have increased, the government set up the working group Anti-Phishing and also implemented several laws for victims [10]. Figure 1 shows phishing sites reported in 2019.

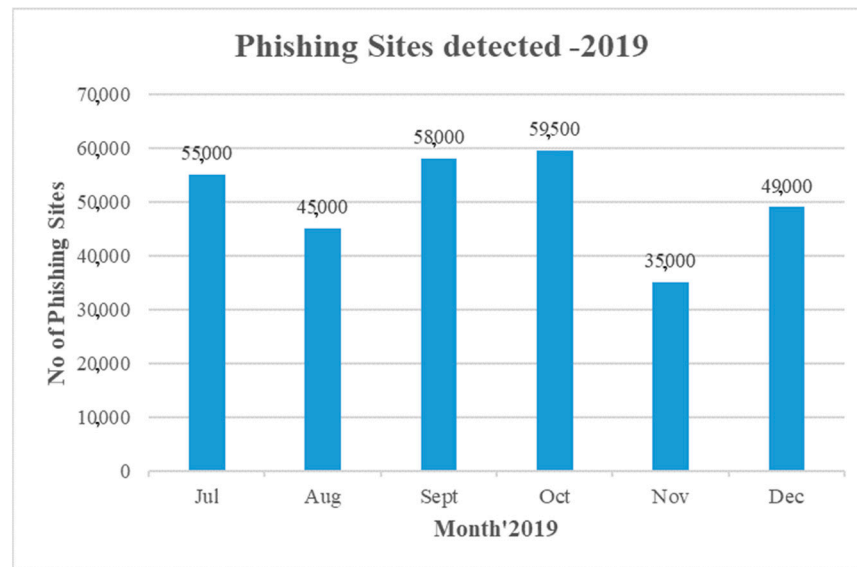


Figure 1. Phishing sites reported in 2019.

1.3. Proposed Research Key Features

The objective of this work is to evaluate the harmfulness of this problem and offers better solutions to protect against phishing attacks.

- Fishing attacks may occur at any time. Thus, based on database features, this work develops a mechanism.
- The proposed work focuses on recent databases and performance can be evaluated based on parameters.
- According to the literature survey, most of the existing work is based on imbalanced mechanisms.
- Three different classifiers are used to determine classification accuracy in detecting phishing.
- The motivation for the current work is the increasing number of phishing attacks; we need to develop a computational automated methodology for detecting phishing.

Figure 2 shows internet usage per year from 2014 to 2019. Figure 3 shows a year-wise online report of phishing attack incidents from 2014 to 2019.

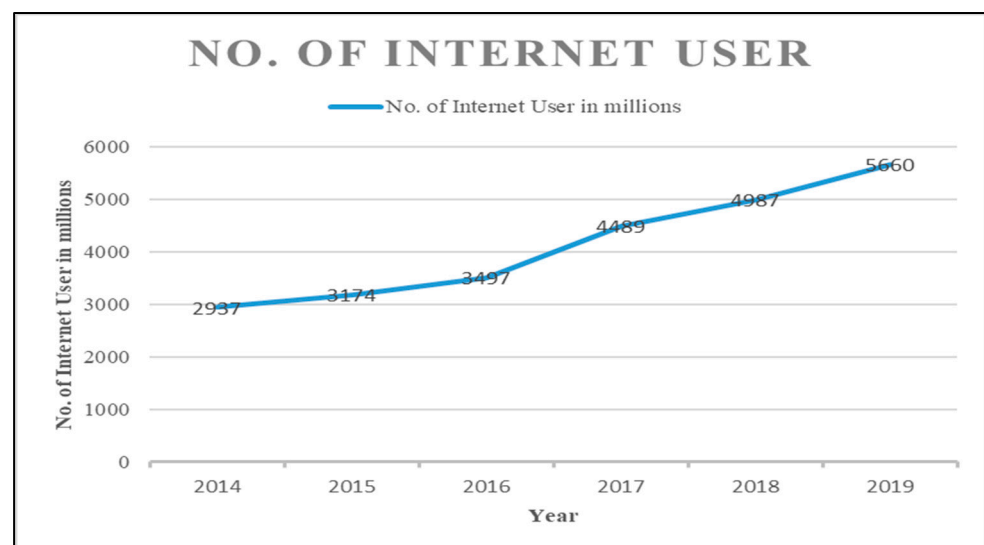


Figure 2. Internet usage per year from 2014 to 2019.



Figure 3. Year-wise online report of phishing attack incidents from 2014 to 2019.

1.4. Motivation

- The motivation for this work is that there is still a lack of awareness regarding phishing threats.
- The main phishing crimes are stealing banking details stealing such as CVV details and credit card information through websites such as PayPal and e-bay.
- Other phishing crimes include theft of personal data and capturing trade secrets and important documents.

2. Literature Survey

A literature survey which is related to the mechanism developed in this investigation is presented in this section. Table 1 summarizes the literature survey for detecting phishing. Figure 4 illustrates phishing attacks in the industry.

Table 1. Literature on Phishing.

Source	Algorithms	Techniques	Merits	Demerits
Babagoli et al. (2018) [11]	Heuristic value based on extracting the parameters	A decision tree and wrapper are used for selecting heuristic-based nonlinear regression.	Using decision trees original dataset can be reduced.	Only phishing and real web pages are used as they contains smaller datasets.
Peng et al. (2018) [12]	Naïve Bayes classifier	A phishing email is identified by using a Naïve Bayes classifier on machine learning and NLP techniques.	To detect the appropriateness of each word, NLP is used.	To establish virus pairs, machine learning is used. Emails text analysis on replying.
Aburrous et al. (2018) [13]	SVM, KNN, Random Forest algorithm, Naïve Bayes	Applying 7 different machine learning processes for the anti-detection process.	Easy to identify the words present in the URL by using NPL features.	To handle large datasets, machine learning will not be more effective.
Kim et al. (2017) [14]	Features of machine learning	Authentication on user and domain levels.	Communication of security can be increased.	The same technology should be used on the sender side and the receiver side.

Table 1. Cont.

Source	Algorithms	Techniques	Merits	Demerits
Zhang et al. (2017) [15]	Neural Networks	A Neural Network was classified with the Monte Carlo algorithm.	Increases accuracy rate and stability detection.	The whole page has to be downloaded.
Drew et al. (2014) [16]	Support Vector Machine	Implementing a transfer of data on prototype between MUA and MTA.	Each sentence can be easily detected by using NLP.	Time consuming and only a small dataset can be used.
Xiang et al. (2011) [17].	CANTINA+	ML methods can be used to identify the phishing site for a content-based system.	Precious work has to be performed to increase the number of specific values. Phishing attacks can be easily understood for evaluation.	Authentication for third-party services. A limited number of datasets used such as 8118 phishing and 4883 original webpages.
Ma et al. (2009) [18]	K-means clustering algorithm	Clustering email, targeted functions, features of 13 orthographic.	Enables achieving reliable results with high effectivity.	Only a K-means algorithm offline technique can be used.
Abu-nimeh et al. (2007) [19]	NNET, BART, Random Forest, CART, Support Vector Machine, Logistic Regression	There is no standard algorithm for detecting phishing by comparing six machine learning algorithms.	To test the classifier, 43 features are used.	Consumes more time and memory if it was used.

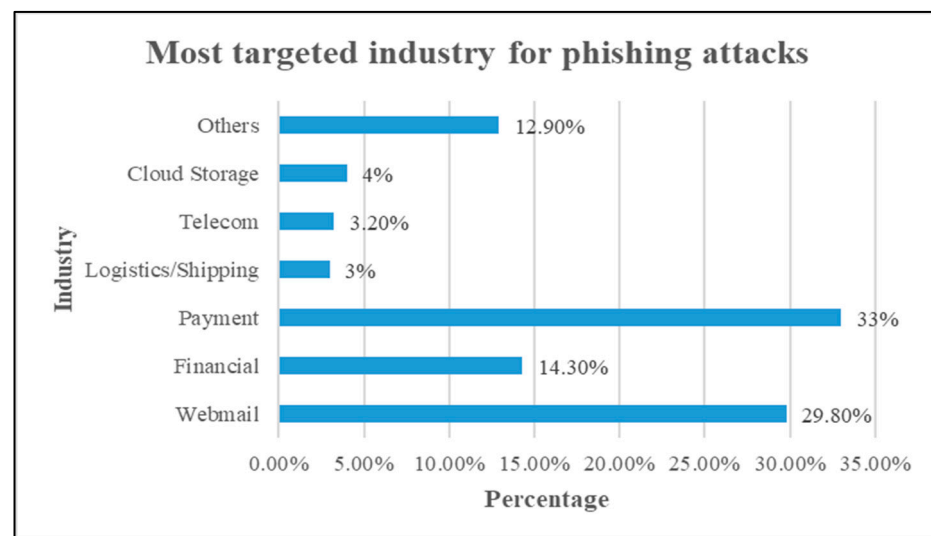


Figure 4. Phishing attacks in industries.

3. The Proposed Methodology

The phishing attack mechanism can be categorized into three categories as

- A DNS blacklist.
- A heuristic-based approach.
- A web crawler-based approach.

These approaches are used for future purposes in phishing attacks for future extraction. Figure 5 shows the proposed architecture of the phishing attack mechanism.

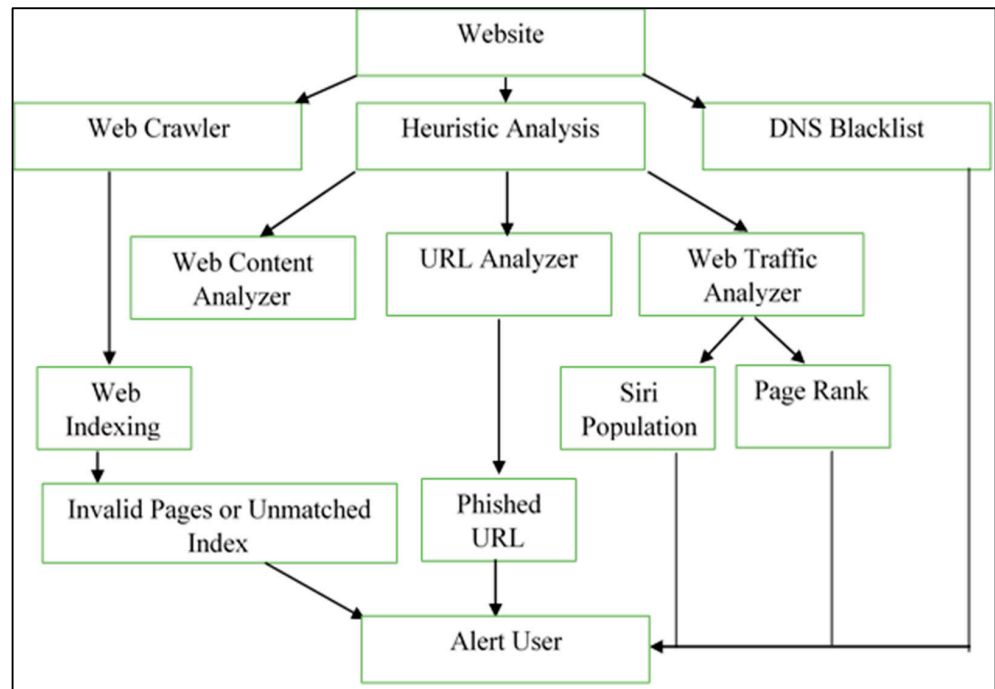


Figure 5. Proposed architecture of the phishing attack mechanism.

3.1. DNS Blacklist and Web Crawler

A DNS blacklist (domain name system blacklist) is used for generating many Internet Protocol addresses which can be easily mounted for programming on the browser. The DNS blacklist is built on the top source file on the internet. This domain name system blacklist generates Internet Protocol addresses with spam purposes. Information is frequently updated on the DNS system. Web crawler starts to attack websites interconnecting with pages and links. Crawling from one website, the phishing attack mechanism goes through all the links in the web index. The proposed phishing attack mechanism crawl is creating a web crawler for each webpage in a website since the attack. Figure 6 shows the crawler for web indexing.

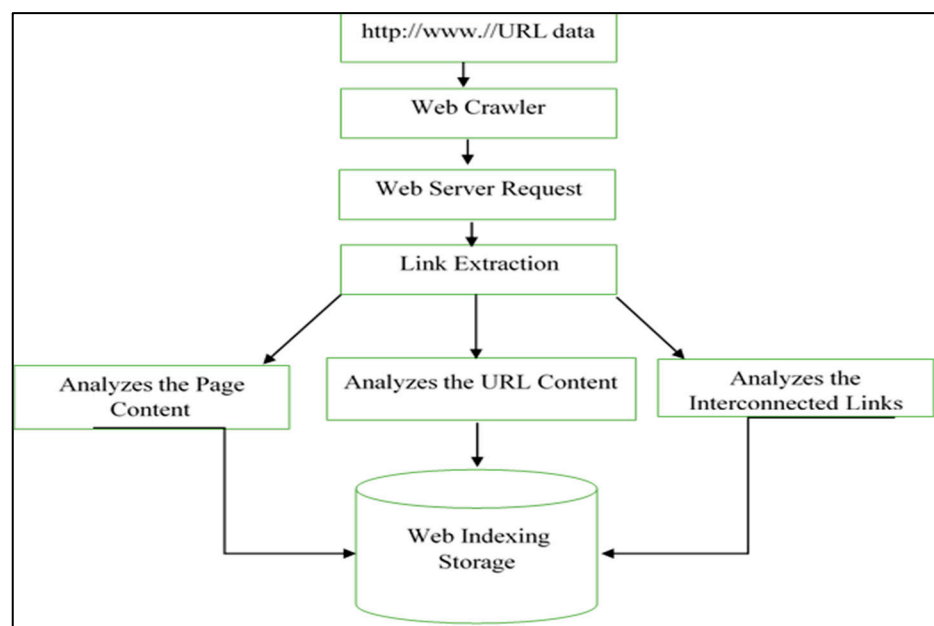


Figure 6. Crawler for web indexing.

3.2. Heuristic Analysis and URL Analysis

Algorithm 1 details the working module of heuristic-based phishing detection. Three features of heuristic analysis phases are as follows. URLs. URL partition is as follows. <protocol>://<subdomain> <primary domain> <TLD>/<path domain>. Algorithm 2 explains the working module of URL-based phishing detection.

Algorithm 1: Heuristic Phishing Detection

Input	Features of URL, content of website, website traffic.
Output	Phishing e-mail or non-phishing site
Step 1	if copyrights are illegal then "User notify" end if
Step 2	(URL initialization → calling algorithm 1 (Web traffic analysis))
Step 3	Counts (total number of counts per visit, number of people visiting per page, visiting duration, bouncing rate)
Step 4	Condition → Operating (PageRank)
Step 5	if Condition is lower then "User notify" end if
Step 6	Condition → Operate (Siri Reputation)
Step 7	if Condition is low then "User notify" end if

Algorithm 2: URL-Based Phishing Detection

Input	Primary domain –k are the features of URL, @, -dots, ID.
Output	Either phished or legal classification.
Step 1	if k is IP address then Condition = Phished else if turned up ('@','-',':');
Step 2	If '@' && '-' Condition = Phished.
Step 3	else if turned up (':')>5 Condition = Phished end if
Step 4	else if ld < 3 Condition = Phished end if
Step 5	if Condition is Phishing then "User notify" end if

3.3. Web Content Analysis and Web Traffic Analyzer

Crawling through the website and web page content copyrights was proposed by the phishing detection mechanism in the website. Regarding suspicion, phishing detection mechanism classifiers send an alert for the message for the contents by the phishing detection mechanism. Parameters are taken from the web traffic analyzer such as tool visits for sites, pages visited per page, duration visiting per person on average, and the bouncing rate. SiriReputation is used on the valued website for calculating the website links from other web pages to itself. PageRank is similar to this [20]. SiriReputation will also be low for phishing the websites in a higher level, for SiriReputation is similar to Pagerank, where SiriReputation values are lower for phishing the websites on a legitimate site.

4. A Machine Learning Approach for Detecting the Attacks

The datasets are collected from Alexa, Siri, and Phish Tank and then processed into machine learning algorithms. This learning algorithm extracts useful information from training examples. The machine learning algorithms can be classified into supervised and semi-supervised. A supervised learning algorithm learns from labelled samples, whereas an unsupervised learning algorithm learns from unknown samples. Initially, the classifier starts with a training phase that can be used to build a decision model. The very important machine learning classifier described here is used for detecting phishing attacks described below. Figure 7 shows the machine learning techniques for detecting phishing attacks.

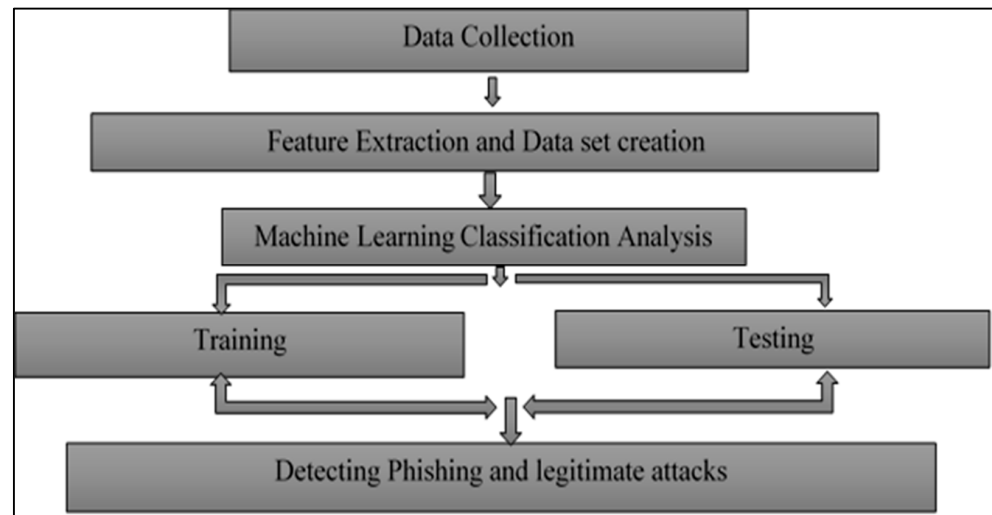


Figure 7. Machine learning techniques for detecting phishing attacks.

4.1. Neural Network Algorithm (NN)

The NN algorithm consists of three layers—an input layer, an output layer and a hidden layer. The hidden layer processes the data and passes it to the output layer. Several attacks are found and recognized by the multi-layer perceptron algorithm. This algorithm is trained by the back-propagation technique, which is based on the concept of feed forward and back propagation [21–23].

4.2. Support Vector Machine (SVM)

The SVM is used for guess and classification, which is used to find the boundaries in multi-dimensional space [24–26]. Its distinct data points can be divided into two classes, +1 and −1, using a hyperplane. Hence, +1 denotes ordinary data and −1 denotes doubtful data.

The hyperplane can be written as Equation (1)

$$WX = b = 0 \quad (1)$$

where $W = w_1, w_2, \dots, w_n$ are weight vectors for n attributes values x_1, x_2, \dots, x_n and b is a scalar. The Support Vector Machine aims to discover the linear best hyperplane so that the boundary of partition between the two classes is magnified. The hyperplane with the peak margin is treated as a good hyperplane. This machine classifies two classes, and multi-class classification is understood by developing an SVM for each two of the classes.

4.3. Random Forest

Random Forests are based on decision trees [27–29]. The computational methodology is the best method to classify phishing in phishing attack mechanisms. Figure 8 shows the extraction parameter in the URL and website. Figure 9 shows the extraction parameter in an e-mail. Table 2 shows the feature extraction and dataset creation.

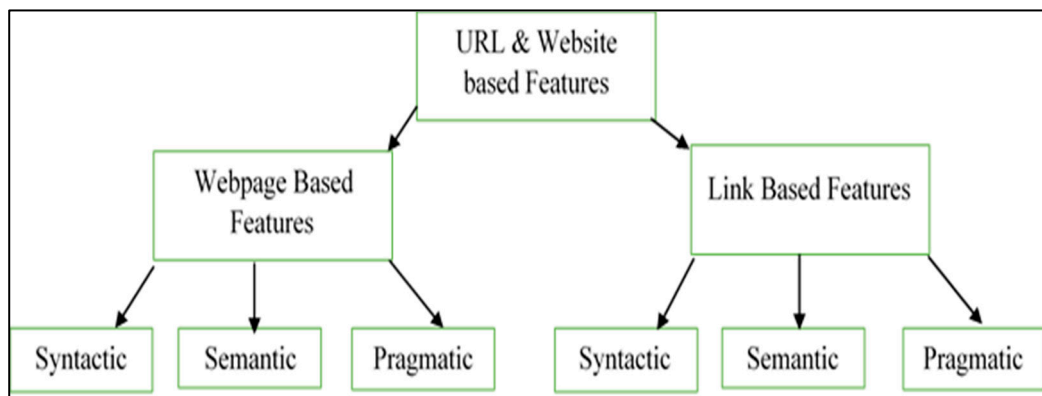


Figure 8. Extraction parameter in the URL and website.

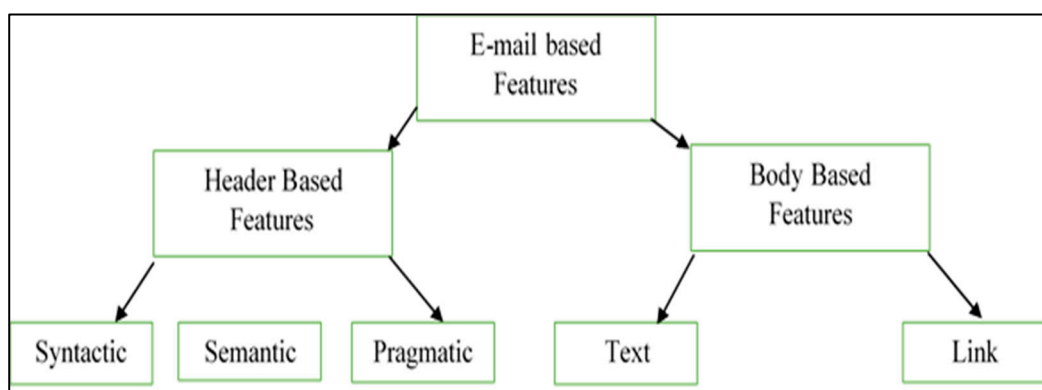


Figure 9. Extraction parameter in e-mail.

Table 2. Feature Extraction and Dataset Creation.

S. No	Feature Extraction	Feature Description
1	Lengthily URL	Websites with a URL length greater than 1750 is likely phishing.
2	Symbol “-”	Domain names including “-” are considered legitimate URLs.
3	URL subdomain	URL subdomains is are likely phishing.
4	HTTPS	This is considered a secure URL.
5	IP address using the domain name	Hackers hide the number with their name when it is phishing.
6	URL request	URLs which consider all images and text together in the same domain.
7	Domain age	Websites created within the last year are likely phishing.
8	Traffic website	Traffic is how many times users have visited a site. A website with no previous history is likely phishing.
9	Domain record	If there is no previous record of the website, it is likely phishing.
10	Pop-up	Pop-up windows are used for hacking, by phishing for passwords.
11	Page redirecting	Redirection to another web page based on a link is likely phishing.
12	URL abnormal	An abnormal URL is identified by the WHO through the domain, based on results identified as phishing.

For training data, 200 samples with features and labels were divided into training and testing learning processes, respectively. Three classification algorithms are used to develop an accurate approach for detecting phishing. The performance of various algorithms was measured by using evaluation metrics on the test samples. A total of 70% of the data were used in the training stage. Testing and validation were processed with the remaining 30% samples.

Sensitivity is described as the ratio of correctly recognized phishing attacks. (1-Specificity) is another attribute of a classifier which describes the ratio of non-phishing attacks. Sensitivity and specificity are the most significant parameters for performance metrics computed by any classifier.

	Predicted Class	
Actual Class	TP	FN
	FP	TN

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$(1 - Specificity) = \frac{FP}{FP + TN}$$

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

$$F1 * Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

where Confusion Matrix is a method to analyze the performance metrics of any classification algorithm and it also provides better solutions.

5. Results

Table 3 shows a comparison of existing work based on features. In general, researchers have achieved an accuracy of approximately 90%. However, in all the previous cases, researchers have restricted their analysis to only one algorithm. Moreover, none of them has considered all areas, i.e., URL, website and email. Thus, the current work has a definite advantage over them in terms of the widespread application area and comprehensive analysis in terms of the several algorithms considered.

Table 3. Comparison of Existing Work Based on Features.

Source	Area	Database	Importance Features	Algorithm	Accuracy
Darling et al. (2015) [10]	URL	PhishTank, Alexa	URL Quadgram, Hostname	Support Vector Machine	90%
Hassan et al. (2017) [30]	Website	UCI Repository based on machine learning	Subdomain, URL Anchor	Neural Network	92%
Toolan et al. (2018) [31]	E-Mail	SpamAssassin Nazario	The number of characters, words, function body and URLs	Support Vector Machine	93%
Lee et al. (2013) [32]	URL	Twitter Applications	Date of account creation, similarity of text by tweeting and friends ratio by followers	Machine learning approach	91%
Rajab et al. (2017) [33].	Website	Yahoo mail directory, PhishTank	Anchor URL, domain name & subdomain name, state of finite SSL	Machine learning approach	94%

Figure 10 illustrates classifier performance. Table 4 depicts the accuracies in % of different classifiers against extracted features. In terms of classifier performance, the

algorithms can be ranked from best to worst as Neural Network > Support Vector Machine > Random Forest. With respect to the Random Forest classifier, the Support Vector Machine and Neural Network classifiers showed 8.32% and 20.65% improvement. A total of 12 parameters were extracted by the algorithms.

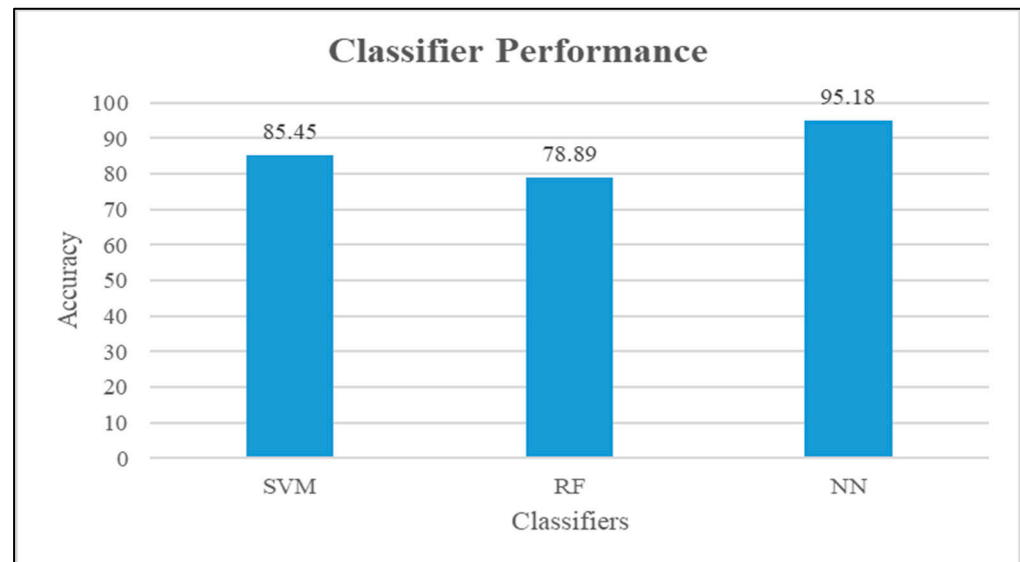


Figure 10. Classifier performance.

Table 4. Accuracies in % of Different Classifiers against Extracted Features.

Support Vector Machine	Machine Learning Accuracy			Total Extracted Parameter
	Random Forest	Neural Network		
85.45%	78.89%	95.18%		12 Parameters

Table 5 shows performance evaluation metrics with different classifiers. In terms of specificity, the Random Forest and Neural Network classifiers are better than the Support Vector Machine classifier by 0.42% and 4.65%, respectively. However, the sensitivity of the Support Vector Machine and Neural Network classifiers is greater than the Random Forest classifier by 9.74% and 21.99%, respectively. Similarly, the precision of the Support Vector Machine and Neural Network classifiers is greater than the Random Forest classifier by 16.46% and 28.77%, respectively. The F1 score of the Support Vector Machine and Neural Network classifiers show 14.27% and 26.66% respective improvements over the Random Forest classifier.

Table 5. Performance Evaluation Metrics with Different Classifiers.

Classifiers	Specificity	F1	Precision	Sensitivity
Support Vector Machine	0.9463	0.8488	0.8757	0.8334
Random Forest	0.9503	0.7428	0.7519	0.7594
Neural Network	0.9903	0.9408	0.9682	0.9264

Figure 11 shows the comparison of different methodologies. The machine learning approach is found to be the best among the three approaches. The heuristic-based approach and the machine learning approach show an improvement of 2.22% and 5.76%, respectively, over the blacklist-based approach.

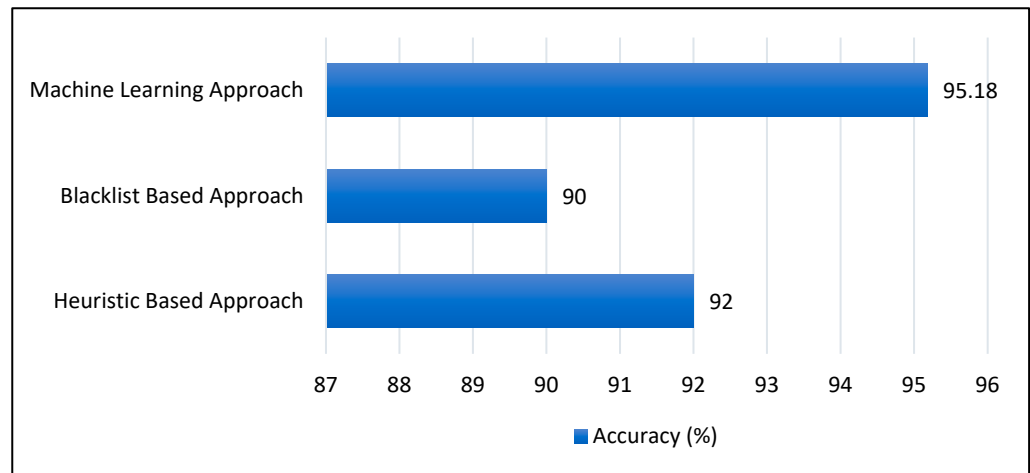


Figure 11. Comparison of different methodologies.

Figure 12 illustrates spam mails detected by the proposed methodology in e-mail. Figure 13 shows fake websites detected and blocked by the proposed methodology in Netcraft. Figure 14 illustrates fake websites detected by the proposed methodology in Google Safe Browsing.

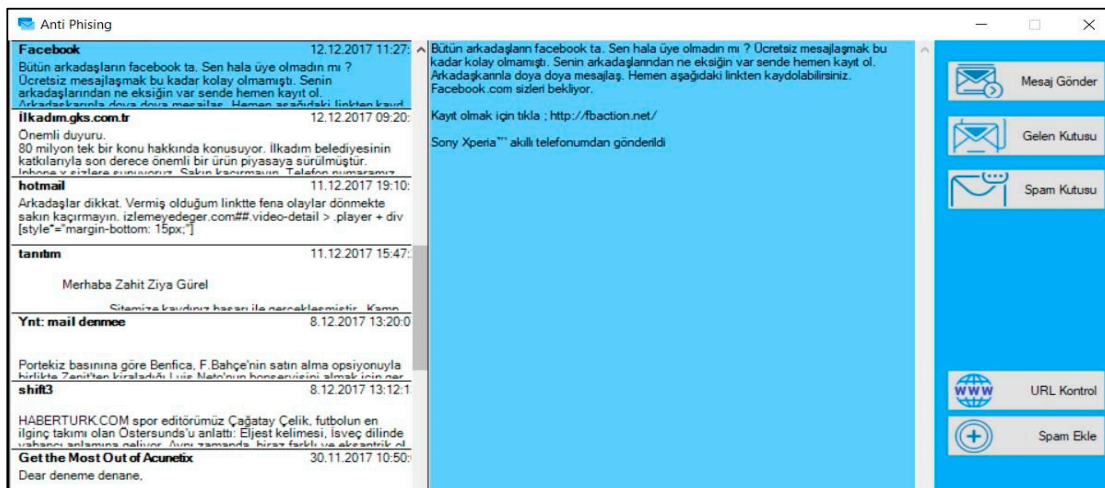


Figure 12. Spam mail detected by the proposed methodology in e-mail.



Figure 13. Fake websites detected and blocked by the proposed methodology in Netcraft.



Figure 14. Fake website detected by the proposed methodology in Google Safe Browsing.

6. Conclusions

A phishing detection mechanism was proposed to detect phishing attackers. The developed phishing detection mechanism is implemented through three phases. Detection based on the DNS blacklist is performed, and then heuristic-based detection is followed by using a web crawler. It is easy to identify the websites frequently using phishing IPs in the DNS blacklist. Using the web crawler and analysis phase, phishing e-mails and sites are identified. The proposed experimental analysis was performed for the phishing detection mechanism and it is used for precisely detecting websites which are phishing as the phishing detection mechanism has the best accuracy. Three different classifiers were used to determine classification accuracy in detecting phishing, resulting in a classification accuracy of 95.18%, 85.45% and 78.89%, for NN, SVM, and RF, respectively. The results suggest that a machine learning approach is best for detecting phishing.

Author Contributions: Conceptualization, G.M., J.V., M.M. and J.A.; data curation, G.M.; formal analysis, G.M.; investigation, G.M.; methodology, J.V. and M.E.; resources, J.V. and M.E.; software, J.V., M.M., J.A. and M.E.; supervision, M.M. and J.A.; writing—original draft, G.M.; writing—review and editing, J.V., M.M., J.A. and M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available through email upon request to the corresponding author.

Acknowledgments: The authors thank M. Ramachandran, lead research scientist, REST Labs, India, for his help in data curation and some of the analyses.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ramanathan, V.; Wechsler, H. phishGILLNET—Phishing detection methodology using probabilistic latent semantic analysis, AdaBoost, and co-training. *EURASIP J. Inf. Secur.* **2012**, *1*, 1–22. [CrossRef]
2. Sophie, G.-P.; Granadillo, G.G.; Laurent, M. Decisive Heuristics to Differentiate Legitimate from Phishing Sites. In Proceedings of the Network and Information Systems Security (SAR-SSI), La Rochelle, France, 18–21 May 2011; IEEE: Manhattan, NY, USA, 2011.
3. Moghimi, M.; Varjani, A.Y. New rule-based phishing detection method. *Expert Syst. Appl.* **2016**, *53*, 231–242. [CrossRef]
4. Prakash, P.; Kumar, M.; Kompella, R.R.; Gupta, M. Phishnet: Predictive Blacklisting to Detect Phishing Attacks. In Proceedings of the 2010 IEEE INFOCOM, San Diego, CA, USA, 14–19 March 2010.
5. Jim, L.; Wong, M. Sender ID: Authenticating E-mail. *RFC 4406*, April 2006.
6. Chen, J.; Guo, C. Online Detection and Prevention of Phishing Attacks. In Proceedings of the 2006 First International Conference on Communications and Networking in China, Beijing, China, 25–27 October 2006; IEEE: Manhattan, NY, USA, 2006; pp. 1–7.

7. Gansterer, W.N.; Polz, D. E-Mail Classification for Phishing Defence. In Proceedings of the 31th ECIR Research on Advances in Information Retrieval, Toulouse, France, 6–9 April 2009; Springer: Cham, Switzerland, 2009; pp. 449–460.
8. Krieg, G.; Kopan, T. CNN News, Is This the Email That Hacked John Podesta’s Account? Available online: <http://edition.cnn.com/2016/10/28/politics/phishing-email-hack-john-podesta-hillary-clinton-wikileaks> (accessed on 19 November 2016).
9. The Trembling Uterus Blog. Available online: <http://tremblinguterus.blogspot.pt> (accessed on 19 November 2016).
10. Darling, M.; Heileman, G.; Gressel, G.; Ashok, A.; Poornachandran, P. A lexical approach for classifying malicious URLs. In Proceedings of the International Conference on High Performance Computing & Simulation (HPCS), Amsterdam, The Netherlands, 20–24 July 2015; pp. 195–202.
11. Babagoli, M.; Aghababa, M.P.; Solouk, V. Heuristic nonlinear regression strategy for detecting phishing websites. *Soft Comput.* **2018**, *23*, 4315–4327. [[CrossRef](#)]
12. Peng, T.; Harris, I.; Sawa, Y. Detecting Phishing Attacks Using Natural Language Processing and Machine Learning. In Proceedings of the IEEE 12th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 31 January–2 February 2018; pp. 300–301.
13. Aburrous, M.; Hossain, M.; Dahal, K.; Thabtah, F. Intelligent phishing detection system for e-banking using fuzzy data mining. *Expert Syst. Appl.* **2010**, *37*, 7913–7921. [[CrossRef](#)]
14. Kim, H.; Lee, E.A. Authentication and Authorization for the Internet of Things. *IT Prof.* **2017**, *19*, 27–33. [[CrossRef](#)]
15. Zhang, Y.; Hong, J.; Cranor, L. Cantina: A content-based approach to detecting phishing websites. In Proceedings of the 16th International World Wide Web Conference (WWW’07), Banff, AB, Canada, 8–12 May 2007; pp. 639–648.
16. Drew, J.; Moore, T. Automatic identification of replicated criminal websites using combined clustering. In Proceedings of the IEEE Symposium on Security and Privacy Workshops (SPW), San Jose, CA, USA, 17–18 May 2014; pp. 116–123.
17. Xiang, G.; Hong, J.; Rose, C.P.; Cranor, L. Cantina+: A feature-rich machine learning framework for detecting phishing web sites. *ACM Trans. Inf. Syst. Secur.* **2011**, *14*, 21. [[CrossRef](#)]
18. Ma, L.; Yearwood, J.; Watters, P. Establishing phishing provenance using orthographic features. In Proceedings of the 2009 eCrime Researchers Summit, Tacoma, WA, USA, 20 September–21 October 2009.
19. Abu-Nimeh, S.; Nappa, D.; Wang, X.; Nair, S. A Comparison of Machine Learning Techniques for Phishing Detection. In Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit 2007, Pittsburgh, PA, USA, 4–5 October 2007; pp. 60–69.
20. Feng, F.; Zhou, Q.; Shen, Z.; Yang, X.; Han, L.; Wang, J. The application of a novel neural network in the detection of phishing websites. *J. Ambient Intell. Humaniz. Comput.* **2018**, 1–15. [[CrossRef](#)]
21. Khonji, M.; Iraqi, Y.; Jones, A. Phishing detection: A literature survey. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 2091–2121. [[CrossRef](#)]
22. Behera, R.R.; Ghadai, R.K.; Kalita, K.; Banerjee, S. Simultaneous prediction of delamination and surface roughness in drilling GFRP composite using ANN. *Int. J. Plast. Technol.* **2016**, *20*, 424–450. [[CrossRef](#)]
23. Magdy, S.; Abouelseoud, Y.; Mikhail, M. Efficient spam and phishing emails filtering based on deep learning. *Comput. Netw.* **2022**, *206*, 108826. [[CrossRef](#)]
24. Ganesh, N.; Joshi, M.; Dutta, P.; Kalita, K. PSO-tuned Support Vector Machine Metamodels for Assessment of Turbulent Flows in Pipe Bends. *Eng. Comput.* **2019**, *37*, 981–1001.
25. Gupta, K.; Kalita, K.; Ghadai, R.; Ramachandran, M.; Gao, X.-Z. Machine Learning-Based Predictive Modelling of Biodiesel Production—A Comparative Perspective. *Energies* **2021**, *14*, 1122. [[CrossRef](#)]
26. Anupam, S.; Kar, A.K. Phishing website detection using support vector machines and nature-inspired optimization algorithms. *Telecommun. Syst.* **2021**, *76*, 17–32. [[CrossRef](#)]
27. Ganesh, N.; Jain, P.; Choudhury, A.; Dutta, P.; Kalita, K.; Barsocchi, P. Random Forest Regression-Based Machine Learning Model for Accurate Estimation of Fluid Flow in Curved Pipes. *Processes* **2021**, *9*, 2095. [[CrossRef](#)]
28. Shanmugasundar, G.; Vanitha, M.; Čep, R.; Kumar, V.; Kalita, K.; Ramachandran, M. A Comparative Study of Linear, Random Forest and AdaBoost Regressions for Modeling Non-Traditional Machining. *Processes* **2021**, *9*, 2015. [[CrossRef](#)]
29. Bhattacharya, S.; Kalita, K.; Čep, R.; Chakraborty, S. A Comparative Analysis on Prediction Performance of Regression Models during Machining of Composite Materials. *Materials* **2021**, *14*, 6689. [[CrossRef](#)]
30. Hassan, D. On determining the most effective subset of features for detecting phishing Websites. *Int. J. Comput. Appl.* **2017**, *122*, 1–7. [[CrossRef](#)]
31. Toolan, F.; Carthy, J. Feature selection for Spam and Phishing detection. In Proceedings of the 2010 eCrime Researchers Summit, Dallas, TX, USA, 18–20 October 2010; pp. 1–12.
32. Lee, S.; Kim, J. WarningBird: A Near Real-Time Detection System for Suspicious URLs in Twitter Stream. *IEEE Trans. Dependable Secur. Comput.* **2013**, *10*, 183–195. [[CrossRef](#)]
33. Rajab, K.D. New Hybrid Features Selection Method: A Case Study on Websites Phishing. *Secur. Commun. Netw.* **2017**, *2017*, 1–10. [[CrossRef](#)]