IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Machine Learning-Based Client-Side Defence Against Web Spoofing Attacks

NAKKA NARASIMHA RAO[1], MUNI TEJASREE[2], DEEPALA LAKSHMI SIVA PAVANI[3], POTLURI MAHESH[4]

#1Assistant Professor in Department of Information Technology, NRI Institute Of Technology, Agiripalli.

#2#3#4 B.Tech with Specialization of Information Technology in NRI Institute Of Technology, Agiripalli.

Abstract: The protection of personal identification numbers and passwords is a significant barrier for cybersecurity. Deceptive login pages soliciting personal information deceive billions of users daily. A variety of nefarious approaches are employed to deceive individuals into accessing harmful websites, such as phishing emails, clickjacking, malware, SQL injection, session hijacking, man-in-the-middle attacks, denial of service, and cross-site scripting. The offender creates a fraudulent yet convincingly comparable website to deceive victims into divulging their credentials. Researchers have offered many security solutions to mitigate these vulnerabilities; however, these methods are both ineffective and susceptible to error. We introduce and implement a client-side defence system that employs machine learning to detect phishing attempts and recognise fraudulent web sites. Our machine learning algorithm serves as a proof of concept for the Google Chrome plugin PhishCatcher, which classifies URLs as either trustworthy or suspicious. The random forest classifier evaluates a login page for authenticity after acquiring four web properties. The precision and validity of the extension were evaluated on multiple real-world web applications. The findings exhibited a precision and accuracy rate of 98.5% when evaluated on 400 authentic URLs and 400 identified phishing URLs. We assessed the latency of our technique using forty phishing URLs. We improved Random Forest by integrating XGBOOST, a technique that evaluates datasets through forest trees or ensembles of estimators to optimise features more efficiently and achieve superior accuracy.

***Index terms** - IOT devices, Support Vector Machine (SVM) and Random Forest (RF).*

1. INTRODUCTION

French research institute INRIA for digital science and technology was the target of phishing attempts in October 2022. A convincing link was sent to French email customers in order to validate their webmail accounts:

https://www.educationonline.nl/Cliquez.ici.cas.inria.fr.cas.login/login.html

The link that was clicked by visitors led them to a sham Inria login page for central authentication, which looked like this: (https://cas.inria.fr/cas/login?service=). The phoney Inria login page looked so identical that users had problems recognising it.

Subscribing users were duped into divulging their Inria credentials by means of a phishing attack. The Inria login page can be accessed by criminals using this sensitive information. It was more probable that consumers would unwittingly provide malicious actors their login credentials due to the similarity between the fake and real sites. In order to decrease phishing threats and increase cybersecurity awareness and vigilance in enterprises, this incident highlights the necessity for comprehensive verification techniques.

Because of technological advancements, e-governance, e-health, online banking, distant education, and e-commerce have all expanded. Social media sites like Twitter and Facebook contribute to the globalisation of the world with their billions of users. Users may sign up for a customised experience on a lot of websites. With a personal account, websites provide certain online services. In the past, users have created accounts on login websites by entering a username and password.

Users will soon be able to access remote resources and services by submitting a web request and then receiving a login form to input their credentials and password. Theft of personal information and identification poses a threat to users' right to privacy. The first step in a phishing assault, as shown in Figure 2, is to send an email with a link to a malicious website [1]. An effort to get the recipient to click on the link might be in the email's content. Clicking on the link leads the unwary individual to believe the website is legitimate. Once the user inputs their credentials and clicks "login," they are sent to the malicious actor. The phisher then uses the credentials to access the legitimate website.

2. LITERATURE SURVEY

a) SpoofCatch: A Client-Side Protection Tool Against Phishing Attacks:

https://www.computer.org/csdl/magazine/it/2021/02/09391742/1sq7EqxybHG

Abstract: The majority of anti-phishing methods found in published works employ intricate traits to identify phishing attempts or evade attack patterns, hence reducing instances of online impersonation. A webpage can allegedly expose a phishing assault, according to this report. By analysing visual similarities across websites, SpoofCatch ensures client protection. For website security, the plugin employs a quartet of similarity algorithms. Prolonged and thorough testing has proven that SpoofCatch can identify and thwart any phishing effort while maintaining a tolerable overhead.

b) Two-Factor Authentication: Too Little, Too Late:

https://profsandhu.com/cs6393_s19/schneier06.pdf

Abstract: It is pointless to not have two-factor authentication. Unfortunately, phishing is still rampant. Identity theft can still occur despite this. Preventing fraud in online accounts is not the objective. It concerns security holes that were present ten years ago. Changing a password is easy. Various forms of encryption were used. Some people write them, while others read them. Hacking into email accounts. Such servers listen in on communications made during remote logins. Additionally, passwords pose a risk. Since the user's identity is now unknown, it can no longer be used as an authentication token. Use two-factor authentication to resolve this issue. Complex numerical passwords or one-of-a-kind answers to challenges created at random are tough to crack. There is no way to write this part because it is always changing. When a password is hacked, it loses all validity. The difficulty of cracking a two-factor password increases. Even trusting a secretary with your token and password has its limitations. Even though these tokens have been around for more than 20 years, their popularity is just now beginning to rise. One of AOL's strengths is distribution. Many financial organisations are already offering these to their customers. Companies adopt two-factor authentication because passwords aren't secure enough.

c) A framework for detection and measurement of phishing attacks:

https://dl.acm.org/doi/10.1145/1314389.131439 1

Abstract: Identity thieves utilise social engineering and advanced attack vectors to acquire sensitive financial information from unsuspecting individuals. A prevalent strategy employed by phishers is to divert consumers to harmful websites. This article examines the creation of phishing URLs. It is generally feasible to identify phishing URLs even in the absence of page data. Phishing URLs can be differentiated from authentic ones by many factors. Integrating these attributes into a model of a dependable logistic regression filter. This filter enables us to assess the prevalence of Internet phishing by analysing several million URLs.

d) ‘An evaluation of machine learning-based methods for detection of phishing sites

https://link.springer.com/chapter/10.1007/978-3-642-02490-0_66

Abstract: In this article, we assess methods for detecting phishing websites that rely on machine learning. We offer nine different machine learning algorithms: AdaBoost, Bagging, Support Vector Machines, Classification and Regression Trees, Logistic Regression, Random Forests, Neural Networks, Naive Bayes, and Bayesian Additive Regression Trees. These machine learning algorithms are trained to identify phishing websites by combining heuristics. We use detection techniques based on machine learning to classify our dataset of 1,500 legitimate and 1,500 phishing sites, and we measure the efficacy of these algorithms. The f 1 measure, error rate, and area under the curve (AUC) were used to evaluate performance with our detection technique requirements. AdaBoost boasts the best AUC (0.9342), lowest error rate (14.15%), and greatest f 1 measure (0.8581). Additionally, seven out of nine detection algorithms based on machine learning outperformed the conventional methods.

e) Detecting phishing websites using machine learning technique

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0258361

In recent years, online commerce and trading have grown due to advancements in cloud computing and the Internet. Organisational resources are harmed and sensitive user data might be accessed unauthorisedly due to growth. The goal of phishing is to obtain sensitive information by tricking victims into visiting malicious websites. Even down to the URL and design, the majority of phishing websites seem just like the actual thing. Many other approaches have been proposed for identifying phishing websites, including heuristics, blacklists, and others. The victims are increasing at an exponential rate because of insufficient security measures. Phishing attacks are more prevalent when users are able to surf the web anonymously. The effectiveness of phishing detection systems is low, according to the available studies. We require a very advanced cyber-defense strategy. Machine learning-based URL detection was introduced in this study. Recurrent neural networks are used to identify malicious URLs. A total of 7,900 malicious and 5,800 legitimate websites were used to test the method. Research shows that compared to existing approaches, the suggested technique for detecting malicious URLs is superior.

3. METHODOLOGY

i) Proposed Work:

Due to the failure of machine learning and signature-based approaches, Random Forest is employed in this study to detect phishing URLs. Random Forest streamlines feature selection and optimisation, leading to enhanced forecast accuracy. To select features and eliminate noise, random forest use tree networks. Primarily, the author includes details in the document. For the purpose of training our URL safety algorithm, we utilised PHISHTANK, a dataset that contains thousands of both legal and illicit URLs.

In addition to Random Forest, the innovative algorithm XGBOOST is useful. It outperforms Random Forest in terms of accuracy and feature optimisation when it comes to filtering datasets using estimators or forest trees. We added XGBOOST.

ii) System Architecture:

The proposed system architecture for PhishCatcher is designed as a client-side defense mechanism that integrates with the Google Chrome browser as a plugin. The architecture incorporates machine learning models to analyze and classify URLs in real-time, identifying them as either trustworthy or suspicious. When a user visits a website, the system extracts four key web properties, such as URL structure, webpage metadata, and embedded features. These properties are then evaluated by a Random Forest classifier, enhanced with the XGBOOST algorithm, which processes the input through forest trees or ensemble estimators to optimize feature analysis and improve classification accuracy. Upon analysis, the plugin provides feedback to the user, either validating the website as safe or flagging it as a potential phishing threat. The system operates efficiently, ensuring minimal latency, and has been validated against a diverse set of authentic and phishing URLs to guarantee its effectiveness in real-world scenarios.

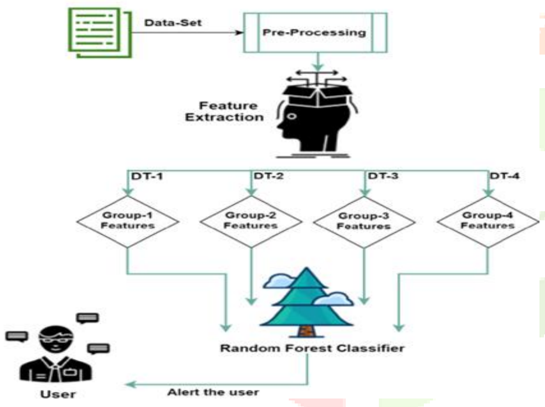
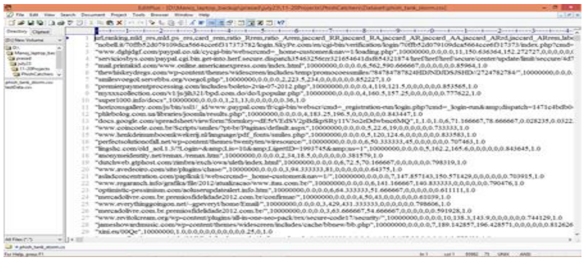


Fig 1 Proposed architecture

iii) Dataset collection:

This collection contains data on problematic behaviours, including cyberbullying that uses racial and ethnic profiling, threats, and hostile language. The data came from Twitter and Facebook groups. Text in tweets and comments is rated as troublesome or not. After data scraping, manually assign -1 and 0 to dubious and non-questionable data. The dataset begins with column names and continues with URLs and labels. Machine learning models will be trained and evaluated using this dataset.


EditPlus - [D:\Manoj_laptop_backup\prasad\july23\11-20Projects\PhishCatchers\Dataset\phish_tank_storm.csv]
File Edit View Search Document Project Tools Browser Window Help
Directory Cliptext
All Files (*.*)
For Help, press F1

```
url,ranking,mld_res,mld.ps_res,card_rem,ratio_Rrem,ratio_Arem,jaccard_RR,jaccard_RA,jaccard_AR,jaccard_AA,jaccard_ARrd,jaccard_ARrem,label
"nobell.it/70ffb52d079109dca5664cce6f317373782/login.SkyPe.com/en/cgi-bin/verification/login/70ffb52d079109dca5664cce6f317373/index.php?cmd=
"www.dghjdgf.com/paypal.co.uk/cycgi-bin/webscrcmd=_home-customer&nav=1/loading.php",10000000,0.0,0.0,11,150.636364,152.272727,0,0,0,0,(
"serviciosbys.com/paypal.cgi.bin.get-into.herf.secure.dispatch35463256rzr321654641dsf654321874/href/href/href/secure/center/update/limit/seccure/4d7
"mail.printakid.com/www.online.americanexpress.com/index.html",10000000,0.0,0.0,6,562,590.666667,0,0,0,0,0,0.85964,1.0
"thewhiskeydregs.com/wp-content/themes/widescreen/includes/temp/promocoessmiles/?84784787824HDJNDJDSJSHD//2724782784/",10000000,0.0,0.
"smilesvoegol.servebbs.org/voegol.php",10000000,0.0,0.0,2,223.5,234,0,0,0,0,0,0.852227,1.0
"premierpaymentprocessing.com/includes/boleto-2via-07-2012.php",10000000,0.0,0.0,4,119,121.5,0,0,0,0,0,0.853565,1.0
"myxxxcollection.com/v1/js/jih321/bpd.com.do/do/l.popular.php",10000000,0.0,0.0,4,160.5,157.25,0,0,0,0,0,0.777622,1.0
"super1000.info/docs",10000000,0.0,0.0,3,21,13,0,0,0,0,0,0.36,1.0
"horizonsgallery.com/js/bin/ssl1/_id/www.paypal.com/fr/cgi-bin/webscr/cmd=_registration-run/login.php?cmd=_login-run&amp;dispatch=1471c4bdb0
"phlebolog.com.ua/libraries/joomla/results.php",10000000,0.0,0.0,4,183.25,196.5,0,0,0,0,0,0.843447,1.0
"docs.google.com/spreadsheet/viewform?formkey=dE5rVEdSV2pBdkpSRy11V3o2dDdwbmc6MQ",1,1.0,1.0,6,71.166667,78.666667,0.028235,0.0322
"www.coincoele.com.br/Scripts/smiles/?pt-br/Paginas/default.aspx",10000000,0.0,0.0,5,22.6,19,0,0,0,0,0,0.733333,1.0
"www.henkdeinumboomkwekerij.nl/language/pdf_fonts/smiles.php",10000000,0.0,0.0,5,120,124.6,0,0,0,0,0,0.833583,1.0
"perfectsolutionofall.net/wp-content/themes/twentyten/wiresource/",10000000,0.0,0.0,6,50.333333,45,0,0,0,0,0,0.707463,1.0
"lingshc.com/old_aol.1.3/?Login=&amp;l.is=10&amp;LigerID=1993745&amp;us=1",10000000,0.0,0.0,5,162.2,165.6,0,0,0,0,0,0.843645,1.0
"anonymeidentity.net/remax./remax.htm",10000000,0.0,0.0,2,34,18.5,0,0,0,0,0,0.381579,1.0
"dutchweb.gtphost.com/zimbra/exch/owa/uleth/index.html",10000000,0.0,0.0,6,72.5,70.166667,0,0,0,0,0,0.798319,1.0
"www.avedeoiro.com/site/plugins/chase/",10000000,0.0,0.0,3,94.333333,81,0,0,0,0,0,0.64375,1.0
"asladconcentration.com/paplkuk1/webscrcmd=_home-customer&nav=1/",10000000,0.0,0.0,7,147.857143,150.571429,0,0,0,0,0,0.703915,1.0
"www.rsgaranch.info/grafika/file/2012/atualizacao/www.itau.com.br/",10000000,0.0,0.0,6,141.166667,140.833333,0,0,0,0,0,0.790476,1.0
"optimistic-pessimism.com/aolusersupdatealert.info.htm",10000000,0.0,0.0,6,64.333333,51.666667,0,0,0,0,0,0.611111,1.0
"mercadolivre.com.br.premiosfidelidade2012.com.br/confirmar/",10000000,0.0,0.0,4,50,45,0,0,0,0,0,0.63039,1.0
"www.everythinggoingon.net/~gpeveryt/home/Email/",10000000,0.0,0.0,3,429,431.333333,0,0,0,0,0,0.798606,1.0
"mercadolivre.com.br.premiosfidelidade2012.com.br/",10000000,0.0,0.0,3,63.666667,54.666667,0,0,0,0,0,0.591928,1.0
"www.revitolcream.org/wp-content/plugins/all-in-one-seo-pack/tex/secure-code17/security/",10000000,0.0,0.0,10,138.3,143.9,0,0,0,0,0,0.744129,1.0
"jameshowardmusic.com/wp-content/themes/widescreen/includes/cache/bbnew/bb.php",10000000,0.0,0.0,7,189.142857,196.428571,0,0,0,0,0,0.812626
"xini.eu/0OQe",10000000,1.0,0.0,0,0,0,0,0,0,0,0,0,0.25,0,1.0
```

Fig 2: Dataset

iv) Data Processing:

Information gleaning from raw data is the result of data processing. Data scientists gather information, sort it, clean it, verify it, analyse it, and then present the results in graphical or textual formats. Manual, mechanical, or electronic data processing is possible. Enhance the usefulness of data and make decisions easier. Businesses are able to enhance their operations and make swift strategic decisions with this. Crucial are advancements in computer software and other forms of automated data processing. Insights for quality management and decision-making can be derived from massive databases, particularly big data.At this point, you should have decided on a dataset from which to extract attributes. Our dataset is comprised of four sources.

v) Feature selection:

Step one of this study was the most challenging and intricate. Another challenge was the scarcity of appropriate datasets. Several writers have suggested anti-phishing strategies based on data mining and machine learning. Unfortunately, the vast majority of training datasets are either unavailable or based on oversimplified assumptions. Research on phishing websites is contentious. This complicates the process of compiling a comprehensive dataset. Notwithstanding this, we meticulously analysed the literature research methods to identify the most effective ones for our model. The most noteworthy method is data set.

vi) Algorithms:

a) **Random Forest (RF):** To get the mean of all the decision trees it trains, Random Forest uses ensemble learning. Because of its large feature set, precision, and ability to handle overfitting, it is effective for intrusion detection.

b) **Support Vector Machine (SVM):** SVM is a powerful supervised learning algorithm used for classification tasks. It creates a hyperplane or a set of hyperplanes in a high-dimensional space to separate different classes. SVM is effective in intrusion detection for its ability to handle complex data relationships and non-linearity.

c) **XGBoost:** Machine learning technique XGBoost (Extreme Gradient Boosting) is efficient and scalable for classification and regression. It implements gradient boosting to generate a sequence of decision trees that repair each other's faults. XGBoost uses regularisation to prevent overfitting, handles missing data, and parallelises tree construction to make it faster and more accurate than gradient boosting. The performance and adaptability of XGBoost make it popular in machine learning competitions and real-world applications, especially in structured/tabular data analysis.

4. EXPERIMENTAL RESULTS

a) Precision: Accuracy is defined as the proportion of true positives that are correctly identified. The formula for precision calculation follows:

Precision = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

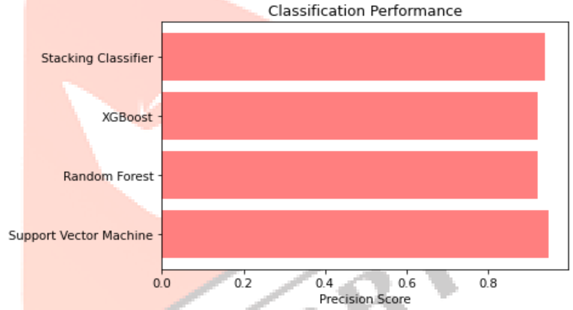


Fig 3 Precision comparison graph

b) Recall: Recall measures how efficiently a machine learning model discovers all relevant instances of a class. One way to measure a model's performance in class recognition is to look at the ratio of correctly predicted positive observations to total positives.

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

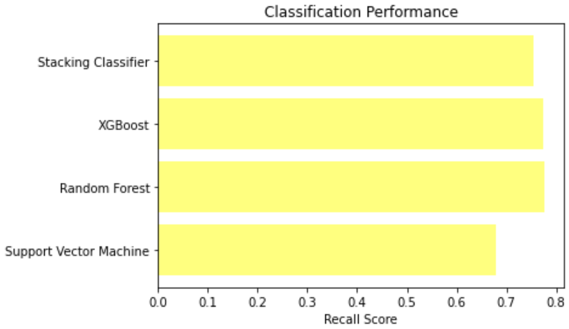


Fig 4 Recall comparison graph

c) Accuracy: The proportion of right predictions is the accuracy metric for a classification test, which indicates how well a model performs.

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

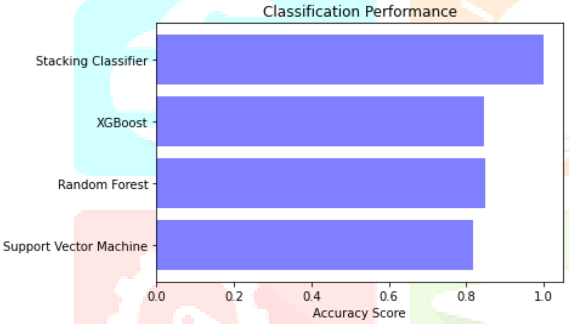


Fig 5 Accuracy graph

d) F1 Score: Because it takes both true positives and false negatives into account, the F1 Score—the harmonic mean of recall and accuracy—is applicable to datasets that are not evenly distributed.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

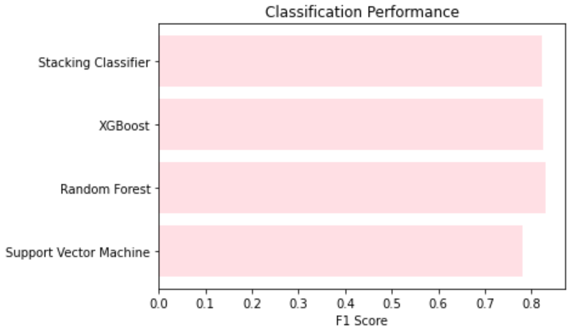
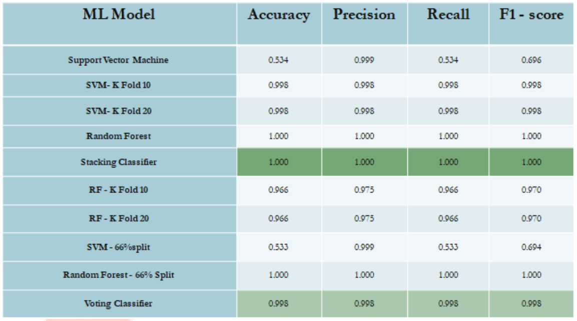
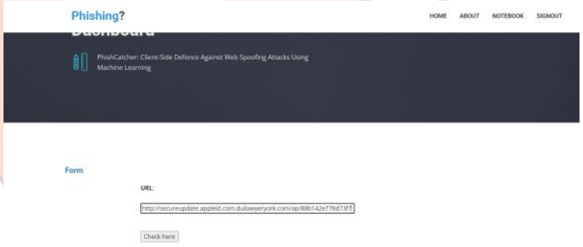


Fig 6 F1Score



ML Model	Accuracy	Precision	Recall	F1 - score
Support Vector Machine	0.534	0.999	0.534	0.696
SVM- K Fold 10	0.998	0.998	0.998	0.998
SVM- K Fold 20	0.998	0.998	0.998	0.998
Random Forest	1.000	1.000	1.000	1.000
Stacking Classifier	1.000	1.000	1.000	1.000
RF - K Fold 10	0.966	0.975	0.966	0.970
RF - K Fold 20	0.966	0.975	0.966	0.970
SVM - 66%split	0.533	0.999	0.533	0.694
Random Forest - 66% Split	1.000	1.000	1.000	1.000
Voting Classifier	0.998	0.998	0.998	0.998

Fig 7 Performance Evaluation

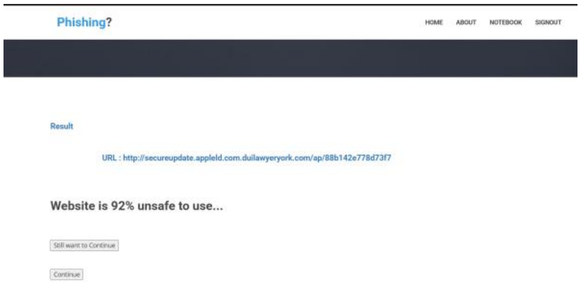


PhishCatcher: Client-Side Defence Against Web Spoofing Attacks Using Machine Learning

Form

URL: http://secureupdate.appleld.com.duilawyeryork.com/ap/88b142e778d73f7

Fig 8 User input



Result

URL : http://secureupdate.appleld.com.duilawyeryork.com/ap/88b142e778d73f7

Website is 92% unsafe to use...

Fig 9 Predict result for given input

5. CONCLUSION

Apps on the web are indispensable because of how well they facilitate online transactions, social networking, virtual education, telemedicine, online banking, digital advertising, and online gaming. Access to restricted online material requires users to register. The dangers to users' privacy and security posed by Internet spoofing have grown. There are a lot of commercial and academic solutions out there to stop online spoofing, but they all have their drawbacks. A basic browser extension that can identify phishing attempts was developed by us using supervised machine learning; we call it PhishCatcher. Unlike other approaches, our technology performs classification within the browser itself. Reducing latency and improving tool efficiency are two ways it helps with online app concerns. Our plug-in includes an easy-to-understand user interface to make everything clear. Phishing URL traits are displayed in a drop-down menu when the user presses a button. Four distinct decision tree groupings are used to categorise the thirty features. In the random forest classifier, decision tree synthesis differentiates between real and bogus login sites. A total of 800 URLs, 400 legitimate and 400 malicious, make up the test set. The True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) confusion matrix is utilised in testing and evaluation. All three metrics for our plug-in—recall, accuracy, and precision—were above 98.5 percent. The plug-in's average latency was 62.5 milliseconds over forty phished URLs. There are thirty features in the collection, however performance could be enhanced with more automated tasks. Use big datasets to train discriminative classifiers like SVM to detect fraudulent URLs. Performance analysis and evaluation techniques can be enhanced with the use of technology.

6. FUTURE SCOPE

Four distinct decision tree groupings are used to categorise the thirty features. Finding fake login websites is a breeze with the random forest classifier since it aggregates decision trees. There are 400 malicious and 400 legitimate URLs in the evaluation and testing collection. True positives, negatives, false positives, and false negatives form the basis of the confusion matrix, which is used to evaluate and test candidates. Thanks to its accurate categorisation, our plug-in stood out. Accuracy and recall are both at 98.5%. When tested with 40 phished URLs, the plug-in had an average latency of 62.5 milliseconds.

Adding automated features could enhance performance, even though the feature set already contains 30 components. By training on larger data, discriminative classifiers like SVM may identify whether URLs are false or legitimate. There are a number of performance analysis tools that help enhance assessment measures.

REFERENCES

[1] W. Khan, A. Ahmad, A. Qamar, M. Kamran, and M. Altaf, ‘‘SpoofCatch: A client-side protection tool against phishing attacks,’’ IT Prof., vol. 23, no. 2, pp. 65–74, Mar. 2021.

[2] B. Schneier, ‘‘Two-factor authentication: Too little, too late,’’ Commun. ACM, vol. 48, no. 4, p. 136, Apr. 2005.

[3] S. Garera, N. Provos, M. Chew, and A. D. Rubin, ‘‘A framework for detection and measurement of phishing attacks,’’ in Proc. ACM Workshop Recurring malcode, Nov. 2007, pp. 1–8.

[4] R. Oppliger and S. Gajek, ‘‘Effective protection against phishing and web spoofing,’’ in Proc. IFIP Int. Conf. Commun. Multimedia Secur. Cham, Switzerland: Springer, 2005, pp. 32–41.

[5] T. Pietraszek and C. V. Berghe, ‘‘Defending against injection attacks through context-sensitive string evaluation,’’ in Proc. Int. Workshop Recent Adv. Intrusion Detection. Cham, Switzerland: Springer, 2005, pp. 124–145.

[6] M. Johns, B. Braun, M. Schrank, and J. Posegga, ‘‘Reliable protection against session fixation attacks,’’ in Proc. ACM Symp. Appl. Comput., 2011, pp. 1531–1537.

[7] M. Bugliesi, S. Calzavara, R. Focardi, and W. Khan, ‘‘Automatic and robust client-side protection for cookie-based sessions,’’ in Proc. Int. Symp. Eng. Secure Softw. Syst. Cham, Switzerland: Springer, 2014, pp. 161–178.

[8] A. Herzberg and A. Gbara, ‘‘Protecting (even naıve) web users from spoofing and phishing attacks,’’ Cryptol. ePrint Arch., Dept. Comput. Sci. Eng., Univ. Connecticut, Storrs, CT, USA, Tech. Rep. 2004/155, 2004.

[9] N. Chou, R. Ledesma, Y. Teraguchi, and J. Mitchell, ‘‘Client-side defense against web-based identity theft,’’ in Proc. NDSS, 2004, 1–16.

[10] B. Hämmerli and R. Sommer, Detection of Intrusions and Malware, and Vulnerability Assessment: 4th International Conference, DIMVA 2007 Lucerne, Switzerland, July 12-13, 2007 Proceedings, vol. 4579. Cham, Switzerland: Springer, 2007.

[11] C. Yue and H. Wang, ‘‘BogusBiter: A transparent protection against phishing attacks,’’ ACM Trans. Internet Technol., vol. 10, no. 2, pp. 1–31, May 2010.

[12] W. Chu, B. B. Zhu, F. Xue, X. Guan, and Z. Cai, ‘‘Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs,’’ in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2013, pp. 1990–1994.

[13] Y. Zhang, J. I. Hong, and L. F. Cranor, ‘‘Cantina: A content-based approach to detecting phishing web sites,’’ in Proc. 16th Int. Conf. World Wide Web, May 2007, pp. 639–648.

[14] D. Miyamoto, H. Hazeyama, and Y. Kadobayashi, ‘‘An evaluation of machine learning-based methods for detection of phishing sites,’’ in Proc. Int. Conf. Neural Inf. Process. Cham, Switzerland: Springer, 2008, pp. 539–546.

[15] E. Medvet, E. Kirda, and C. Kruegel, ‘‘Visual-similarity-based phishing detection,’’ in Proc. 4th Int. Conf. Secur. privacy Commun. Netowrks, Sep. 2008, pp. 1–6.

[16] W. Zhang, H. Lu, B. Xu, and H. Yang, ‘‘Web phishing detection based on page spatial layout similarity,’’ Informatica, vol. 37, no. 3, pp. 1–14, 2013.

[17] J. Ni, Y. Cai, G. Tang, and Y. Xie, ‘‘Collaborative filtering recommendation algorithm based on TF-IDF and user characteristics,’’ Appl. Sci., vol. 11, no. 20, p. 9554, Oct. 2021.

[18] W. Liu, X. Deng, G. Huang, and A. Y. Fu, ‘‘An antiphishing strategy based on visual similarity assessment,’’ IEEE Internet Comput., vol. 10, no. 2, pp. 58–65, Mar. 2006.

[19] A. Rusu and V. Govindaraju, ‘‘Visual CAPTCHA with handwritten image analysis,’’ in Proc. Int. Workshop Human Interact. Proofs. Berlin, Germany: Springer, 2005, pp. 42–52.

Author profile:

Mr. NAKKA NARASIMHA RAO is presently working as Assistant Professor in the department of Information Technology at NRI Institute of Technology, Vijayawada. He received his M.Tech degree from Jawaharlal Nehru Technological University, Kakinada (JNTUK). He has published over 2 research paper in international journals. He has more than 7 years of experience in teaching.

MUNI TEJASREE B.Tech student with Specialization of Information Technology in NRI Institute Of Technology, With a strong interest in technology, I aim to solve real-world problems and have hands-on experience in Java Full Stack Development, Python, and frontend technologies like HTML and JavaScript. Additionally, I have experience with DevOps tools, particularly Jenkins and AWS, which I used to automate CI/CD processes during my internship at Advaita Global - IT Labs Pvt. Ltd. I have also completed NPTEL certifications on "The Joy of Computing Using Python" and "Cloud Computing", enhancing my proficiency in Python programming and cloud technologies. I am passionate about problem-solving and DSA, showcasing my aptitude for tackling coding challenges effectively.

DEEPALA LAKSHMI SIVA PAVANI B.Tech student with Specialization of Information Technology in NRI Institute of Technology, with a strong interest in Image Processing, Data Visualization, and Kubernetes. In addition to academic studies, she has completed the NPTEL certifications on "The Joy of Computing Using Python" and "Cloud Computing," which have enhanced her proficiency in Python programming and cloud technologies. She also earned a Java Full Stack certification from Wipro, gaining knowledge in front-end and back-end development, along with experience in frameworks such as Spring and Hibernate, as well as database management. Through internships at Advaita Global and BIST Technologies, she has gained hands-on experience in Kubernetes deployments, image processing, and data science projects.

POTLURI MAHESH B.Tech student with Specialization of Information Technology in NRI Institute of Technology, with a strong interest in Machine Learning, Data Science, and Deep Learning. In addition to the academic studies he had completed the NPTEL certification on "Joy of Computing using Python", enhancing their proficiency in Python programming and its applications. He had earned a Java Full Stack certification from Wipro, gained knowledge in both front-end and back-end development.