



A HYBRID MACHINE LEARNING SYSTEM FOR URL-BASED PHISHING DETECTION

Punna Lakshmi lavanya , Mr.N.Naveen Kumar

M tech(Computer Science),Student, Department of Information Technology, Associate Professor of CSE, JNTUHUCESTH, Hyderabad,Telangana – 500085

Abstract - I offer a phishing URL-based dataset including phishing and lawful URL credits from more than 11000 site datasets. Phishing URLs are forestalled and clients safeguarded by many ML approaches subsequent to preprocessing. Decision trees, linear regression, random forests, naive Bayes, gradient boosting classifiers, support vector classifiers, and a hybrid LSD model that combines decision trees, SVMs, and logistic regression with soft and hard voting are utilized to really and accurately defend against phishing attacks. The LSD model purposes matrix search hyper boundary improvement and shade include choice with cross overlap approval to make a half breed model by consolidating forecasts from numerous models like Stacking Classifier, an ensemble technique, and Random Forest and MLP Classifier as base classifiers. It utilizes LGBM Classifier as a meta-assessor to figure the end-product, further developing characterization execution. To outline the models' effects and viability, the proposed method was evaluated utilizing a few measures. Precision, accuracy, recall, F1-score, and explicitness were thought of. Examination investigation shows that the proposed strategy outflanks different models and yields the best outcomes.

Keywords:- Phishing attacks, Machine learning algorithms, Cyber threat detection, Hybrid LSD model, Cyber security measures

I. INTRODUCTION

Phishing is a sneaky online threat where cybercriminals impersonate trustworthy sources, like banks or popular websites, to trick individuals into revealing sensitive information such as passwords, credit card numbers, or personal details. Detecting phishing attempts is crucial because it helps prevent valuable data from falling into the wrong hands and protects against financial losses. Machine Learning, a type of artificial intelligence, is highly effective in the fight against phishing. It works by examining large volumes of data, learning patterns from it, and using this knowledge to identify phishing attempts. One significant advantage is that ML systems can adapt to new and evolving phishing techniques, making them very robust. One way to detect phishing is by analyzing website addresses or URLs. Phishers often make mistakes in URLs, like using misspelled domain names or adding too many subdomains. Machine Learning models excel at spotting these subtle irregularities. nEffective phishing detection systems



can be seamlessly integrated into various online tools such as web browsers, email clients, or corporate networks. These integrated systems work in real-time, continuously scanning incoming data for potential phishing threats and providing immediate protection to users.

In this technological era, the Internet has made its way to become an inevitable part of our lives. It leads to many convenient experiences in our lives regarding communication, entertainment, education, shopping and so on. As we progress into online life, criminals view the Internet as an opportunity to transfer their physical crimes into a virtual environment. The Internet not only provides convenience in various aspects but also has its downsides, for example, the anonymity that the Internet provides to its users.[7] With the exponential growth of Internet users, incidents of cybercrimes are also correspondingly expanding in a rapid way. Both people and associations are losing millions worth daily (Hong, 2012, Ragucci and Robila, 2006, University of Portsmouth, 2016). Phishing is one of the basic cybercrimes, which is exponentially increasing day by day.[12] With the rise of the internet era, malicious actors have also been increasing in number. Phishing attacks became a trend in the age where websites are part of everyday life. The exploitation of human weaknesses is a major factor in the victimization of users. Phishing websites are set up to look legitimate or similar to other well established websites, causing the victims of the scam to fall prey to it. Since the malicious sites are sometimes indistinguishable from the legitimate source, nonprofessional users of the internet cannot distinguish between the two. This led

to the creation of phishing blacklists. Phishing blacklists are software datasets that are kept by professionals. They allow nonprofessional users to become aware of potential phishing websites that they might navigate to.[18]

II. LITERATURE SURVEY

Y. Lin, R. Liu, D. M. Divakaran, J. Y. Ng, Q. Z. Chan, Y. Lu, Y. Si, F. Zhang, and J. S. Dong present "Phishpedia," a pioneering logo-based phishing identification system characterized by exceptional accuracy and minimal runtime impact. This innovative deep learning system excels in precise phishing identification, particularly in logo recognition and matching, surpassing current methods. Its proficiency not only outperforms existing techniques but also uncovers previously unidentified phishing sites, thereby fortifying defense against phishing attacks. Phishpedia stands out as a unique and powerful tool for enhancing cybersecurity. *Cons:* Phishpedia's performance relies on logo availability and quality on webpages. Ongoing updates and maintenance are essential for adapting to evolving phishing tactics.[1]

Shirazi, Haynes, and Raya present a pioneering mobile-friendly phishing detection algorithm leveraging Artificial Neural Networks (ANNs) to scrutinize URL and HTML features. Their approach integrates cutting-edge deep transformers such as BERT, ELECTRA, RoBERTa, and MobileBERT for efficient learning from URL text. The innovative system facilitates swift training, seamless maintenance, and real-time deployment on mobile devices, addressing mobile security challenges



effectively. This ensures competitive performance, establishing a robust defense against phishing threats while optimizing resource utilization for enhanced cybersecurity on mobile platforms. *Cons:* Limited to URL detection may miss complex phishing within legitimate pages. Depends on pre-trained transformers, subject to variations in availability and quality. [2]

The thesis by A. Akanchha delves into the realm of SSL certificates within phishing sites, scrutinizing attacker attributes and crafting an auto-detection system reliant on SSL certificate features. Embracing Decision Tree [4] machine learning for its transparency and efficacy, the research presents a pioneering SSL certificate-based phishing detection system, boasting impressive accuracy and a user-friendly Web API. The work underscores the need for future adaptations to combat evolving phishing techniques and ensure ongoing system updates, providing a comprehensive approach to cybersecurity challenges. *Cons:* The system's effectiveness relies on SSL certificate attributes, which could be undermined if attackers develop new methods to mimic genuine certificates. The scalability of the system for managing numerous domains is not extensively discussed.[3]

In the collaborative work of H. Shahriar and S. Nimmagadda, their chapter focuses on Network Intrusion Detection Systems (IDS) leveraging machine learning techniques such as Gaussian Naive Bayes, logistic regression, Decision Tree [4], and neural networks. The study aims to discern normal and anomalous network activities, particularly across TCP/IP layers. Notably, the Decision Tree [4] exhibits

commendable performance on public datasets, yet the authors underscore the imperative of real-world testing and scalability assessments for comprehensive validation of its accuracy and efficiency in practical network intrusion detection scenarios. *Cons:* Evaluation may not reflect real-world conditions or evolving attacks. Algorithm choice not exhaustive; different methods may yield different results.[4]

A. K. Dutta's innovative approach utilizes Random Forest [4], a supervised machine learning technique, to construct an advanced system dedicated to identifying phishing websites. The method involves meticulous analysis and selection of pertinent features that distinctly define phishing sites. Implemented as an intelligent browser extension, the system achieves an impressive 98.8% accuracy in detecting phishing sites, strategically addressing human vulnerabilities in online security. While occasionally presenting false alerts, the overarching goal is to significantly enhance online security measures and provide users with a robust defense against potential cyber threats. *Cons:* Feature quality impacts adaptability to new phishing tactics. Potential for false results affects user trust. [5]

III. METHODOLOGY

A) System Architecture

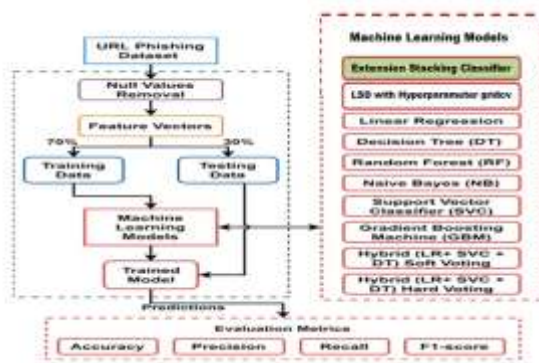


Fig 1: System Architecture

Proposed work

The proposed study depends on a phishing URL-based dataset got from a notable dataset vault, which incorporates phishing and lawful URL credits removed from more than 11000 site datasets. Following preprocessing, a few machine learning calculations have been created and executed to forestall phishing. To really and accurately guard against phishing attacks, this review utilizes ML models like decision trees, linear regression, random forests, naive Bayes, gradient boosting classifiers, support vector classifiers, and proposed hybrid LSD models that combine decision trees, support vector machines, and logistic regression with both soft and hard voting, and applied a hybrid model by joining the expectations of numerous singular models like S. The last prediction is made involving the LGBM Classifier as a meta-assessor, which grows the venture's opportunities for better characterization performance. The proposed LSD model utilizes the overhang feature selection methodology related to cross-fold validation and framework search hyper boundary improvement strategies.

B) Dataset Collection

The "URL-based phishing dataset" is a collection of data designed for the purpose of studying and developing systems to detect and differentiate between phishing and legitimate URLs. It was sourced from Kaggle, a popular platform for data science competitions and datasets.

Here is a general description of the dataset:

Name: URL-based Phishing Dataset

Source: Kaggle

Purpose: To facilitate research and development of phishing detection systems.

Size: Contains data from over 11,000 websites.

Format: Presented in vector form, implying that each URL is likely represented as a set of features or attributes.

The dataset is likely structured in a way that each entry or instance corresponds to a URL, and the features (vector form) associated with each URL provide information that machine learning models can use to make predictions about whether a given URL is associated with phishing or is legitimate.

Typical features in a phishing detection dataset might include characteristics such as the length of the URL, the presence of certain keywords, the use of HTTPS, domain age, and other relevant indicators. These features are crucial for training machine learning models to discern patterns that can differentiate between legitimate and phishing URLs.

url	length	domain	ip_address	is_phishing
http://www.example.com	25	www	192.168.1.1	0
http://www.phishing.com	25	www	192.168.1.1	1
http://www.example.com	25	www	192.168.1.1	0
http://www.phishing.com	25	www	192.168.1.1	1



C) Pre-processing

Using Pandas Data frame: In this step, we leverage Pandas, a powerful data manipulation library in Python, to clean, transform, and prepare the dataset. This involves handling missing values, converting data types, and structuring the data for further analysis or modeling.

Visualization with Seaborn & Matplotlib: Utilizing Seaborn and Matplotlib, we create visualizations such as charts and graphs to gain insights into the dataset's characteristics. This step helps us understand patterns, relationships, and distributions within the data, aiding in informed decision-making for subsequent analysis.

Label Processing: Here, we employ a label encoder, a preprocessing technique, to convert categorical labels into numerical values. This is crucial for machine learning models, as they typically require numerical inputs. Label processing ensures that the models can effectively interpret and learn from the categorical information present in the dataset.

Feature Selection: In this step, we identify and select the most relevant features from the dataset. Feature selection is vital for improving model performance by focusing on the most informative variables and reducing noise. Techniques such as statistical tests, correlation analysis, or machine learning algorithms can be applied to identify the features that contribute significantly to the predictive power of the model.

D) Training & Testing

In the initial phase of our project, we implemented the first machine learning model (Model 9) to analyze and interpret the preprocessed dataset. Following this, during the extension phase, we sought to enhance

predictive accuracy by creating a hybrid model that amalgamates predictions from multiple models. This innovative approach aims to leverage the strengths of diverse models, fostering improved overall accuracy in our predictions. Simultaneously, we developed a user-friendly Flask-based frontend, fortified with authentication measures, to streamline user interaction with the models. This frontend provides a seamless interface for users to input data and obtain predictions, ensuring a practical and accessible experience. The heart of our project lies in training the aforementioned machine learning models on the preprocessed dataset, allowing them to discern intricate data patterns and relationships. Following the training phase, rigorous evaluations are conducted on a distinct test dataset. Performance metrics such as accuracy, precision, recall, and F1-score are meticulously employed to assess the effectiveness of these models in detecting phishing URLs. This robust evaluation process serves as a crucial quality assurance step, ensuring that the models not only exhibit accuracy but also reliability, affirming their suitability for real-world applications. Through this comprehensive methodology, our project aims to deliver advanced and trustworthy solutions in the realm of phishing URL detection.

E) Algorithms.

LR: Logit models are utilized for order and prescient examination. In light of autonomous factors, logistic regression computes the probability of an event, like voting or not voting.

RF: The Leo Breiman and Adele Cutler-licensed random forest technique blends the result of various choice trees to deliver a solitary result. Since it settles



characterization and relapse issues, its straightforwardness and adaptability have made it well known.

Decision trees: Decision trees are non-parametric supervised learning calculations utilized for classification and regression. Its tree structure contains a root hub, branches, inside hubs, and leaf hubs.

SVM: This solid supervised procedure functions admirably on more limited yet muddled datasets. SVMs might be utilized for regression and classification, despite the fact that they perform better in grouping.

Naive Bayes: Naive Bayes methods are supervised learning calculations that utilize Bayes' hypothesis with the "naive" suspicion of restrictive freedom between each sets of qualities given the class variable.

Gradient Boosting: Gradient Boosting is an unmistakable ML approach for classification and regression. Ensemble Learning techniques like boosting train models continuously and attempt to address one another. Consolidating powerless students makes areas of strength for them.

Hybrid LSD Soft: This ML strategy joins the qualities of the Locally Sensitive Discriminant analysis (LSD) and Soft algorithms. The LSD calculation tracks down designs, though the Soft algorithm handles commotion. In this manner, the Hybrid LSD Soft technique can find designs in uproarious information, making it an important information examination device.

Hybrid LSD Hard: This streamlining strategy joins the qualities of the Locally Sensitive Discriminant (LSD) and Hard calculations. The LSD calculation tracks down designs in information, while the Hard strategy

handles commotion free information. These two strategies permit the Hybrid LSD Hard calculation to successfully find designs in commotion free information, making it a valuable information examination device.

LSD with Hyperparameter grid cv: Locally Sensitive Discriminant Analysis model selection utilizing Hyperparameter matrix cv. It attempts a few boundaries and cross-approves to get the ideal blend. LSD works better, working on model accuracy and generalizability.

Stacking Classifier: An ensemble learning approach called a stacking classifier joins various characterization models into one "super" model. This can support execution since the joined model can gain from each model's abilities.

IV. EXPERIMENTAL RESULTS

A) Comparison Graphs → Accuracy, Precision, Recall, f1 score

Accuracy: A test's accuracy is defined as its ability to recognize debilitated and solid examples precisely. To quantify a test's exactness, we should register the negligible part of genuine positive and genuine adverse outcomes in completely examined cases. This might be communicated numerically as:

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



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

examples of a particular class. It is the proportion of appropriately anticipated positive perceptions to add up to real up-sides, which gives data about a model's capacity to catch instances of a specific class.

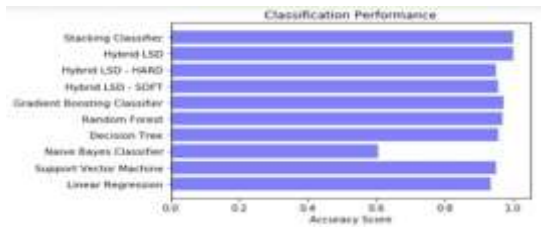


Fig 2: Accuracy Graph

Precision: Precision measures the proportion of properly categorized occurrences or samples among the positives. As a result, the accuracy may be calculated using the following formula:

$$\text{Precision} = \frac{\text{True positives}}{(\text{True positives} + \text{False positives})} = \frac{TP}{TP + FP}$$

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

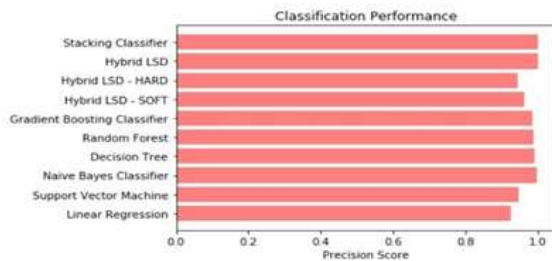


Fig 3: Precision Score Graph

Recall: Recall is a machine learning metric that surveys a model's capacity to recognize all pertinent

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

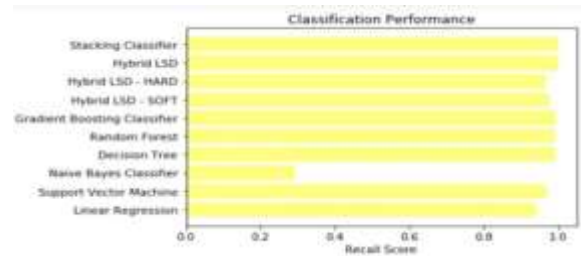


Fig 4: Recall Score Graph

F1-Score: The F1 score is a machine learning evaluation measurement that evaluates the precision of a model. It consolidates a model's precision and review scores. The precision measurement computes how often a model anticipated accurately over the full dataset.

$$\text{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

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

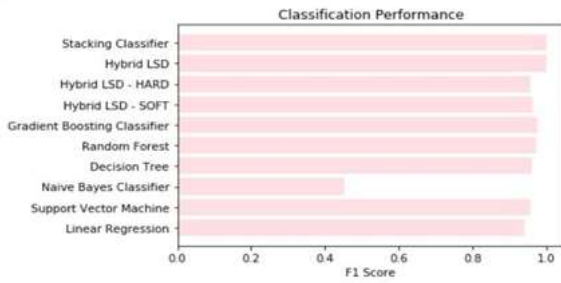


Fig 5: F1 Score Graph

B) Performance Evaluation table.

	ML Model	Accuracy	F1_score	Recall	Precision	Specificity
0	Linear Regression	0.934	0.941	0.943	0.927	0.909
1	Support Vector Machine	0.951	0.957	0.969	0.947	0.909
2	Naive Bayes Classifier	0.905	0.454	0.292	0.197	0.908
3	Decision Tree	0.967	0.962	0.991	0.993	0.909
4	Random Forest	0.969	0.972	0.993	0.991	0.909
5	Gradient Boosting Classifier	0.974	0.977	0.994	0.988	0.909
6	Hybrid LSD - SOFT	0.959	0.954	0.972	0.965	0.909
7	Hybrid LSD - HARD	0.950	0.956	0.987	0.945	0.909
8	Hybrid LSD	1.000	1.000	1.000	1.000	0.426
9	Stacking Classifier	1.000	1.000	1.000	1.000	0.426

Fig 6: Performance Evaluation Table



Fig 7: Home page



Fig 8: User Signup page



Fig 9: User Sign in Page



Fig 10: Enter URL



Fig 11: Url result unsafe 100%





Fig 12: inbioma screen



Fig 13: Url result safe 98%



Fig 14: Amazon website page

V. CONCLUSION

The exploration shows that phishing attacks are a serious and destructive cybercrime on the web, and there is as of now no complete and successful strategy to battle them. Machine learning is basic in fighting cybercrime, especially phishing endeavors. Thus, this recommended strategy utilizes an assortment of ML models, for example, decision trees, linear regression, random forests, naive Bayes, gradient boosting classifiers, support vector classifiers, and a proposed hybrid LSD model. The LSD model combines logistic regression, support vector machine, and decision tree with soft and hard voting. So, the proposed phishing detection system considering a hybrid ML approach

outflanks different models and accomplishes the best outcomes. We likewise utilized a hybrid model by joining the forecasts of different individual models, like the Stacking Classifier, a ensemble strategy, to consolidate expectations from the Random Forest Classifier and the MLP Classifier as base classifiers. The last expectation is made involving the LGBM Classifier as a meta-assessor, which extends the venture's opportunities for better classification execution.

VI. FUTURE SCOPE

The future degree incorporates fostering a constant phishing detection system, investigating the utilization of DL procedures for phishing detection, trying different things with different highlights for phishing detection, assessing the proposed approach on a bigger and more assorted dataset, and making an easy-to-understand interface for the proposed framework. The venture can develop by examining new calculations and procedures to improve phishing detection accuracy and effectiveness. Stretch out the drive to integrate ongoing checking of arising digital dangers.

REFERENCES

- [1] Y. Lin, R. Liu, D. M. Divakaran, J. Y. Ng, Q. Z. Chan, Y. Lu, Y. Si, F. Zhang, and J. S. Dong, "Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages," in Proc. 30th USENIX Secur. Symp. (USENIX Security), 2021, pp. 3793–3810.
- [2] H. Shirazia, K. Haynesb, and I. Raya, "Towards performance of NLP transformers on URL-based



- phishing detection for mobile devices,” Int. Assoc. Sharing Knowl. Sustainability (IASKS), Tech. Rep., 2022.
- [3] A. Akanchha, “Exploring a robust machine learning classifier for detecting phishing domains using SSL certificates,” Fac. Comput. Sci., Dalhousie Univ., Halifax, NS, Canada, Tech. Rep. 10222/78875, 2020.
- [4] H. Shahriar and S. Nimmagadda, “Network intrusion detection for TCP/IP packets with machine learning techniques,” in Machine Intelligence and Big Data Analytics for Cybersecurity Applications. Cham, Switzerland: Springer, 2020, pp. 231–247.
- [5] A. K. Dutta, “Detecting phishing websites using machine learning technique,” PLoS ONE, vol. 16, no. 10, Oct. 2021, Art. no. e0258361.
- [6] A. K. Murthy and Suresha, “XML URL classification based on their semantic structure orientation for web mining applications,” Proc. Comput. Sci., vol. 46, pp. 143–150, Jan. 2015.
- [7] A. A. Ubing, S. Kamilia, A. Abdullah, N. Jhanjhi, and M. Supramaniam, “Phishing website detection: An improved accuracy through feature selection and ensemble learning,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 1, pp. 252–257, 2019.
- [8] A. Aggarwal, A. Rajadesingan, and P. Kumaraguru, “PhishAri: Automatic realtime phishing detection on Twitter,” in Proc. eCrime Res. Summit, Oct. 2012, pp. 1–12.
- [9] S. N. Foley, D. Gollmann, and E. Snekkenes, Computer Security— ESORICS 2017, vol. 10492. Oslo, Norway: Springer, Sep. 2017.
- [10] P. George and P. Vinod, “Composite email features for spam identification,” in Cyber Security. Singapore: Springer, 2018, pp. 281–289.
- [11] H. S. Hota, A. K. Shrivastava, and R. Hota, “An ensemble model for detecting phishing attack with proposed remove-replace feature selection technique,” Proc. Comput. Sci., vol. 132, pp. 900–907, Jan. 2018.
- [12] G. Sonowal and K. S. Kuppasamy, “PhiDMA— A phishing detection model with multi-filter approach,” J. King Saud Univ., Comput. Inf. Sci., vol. 32, no. 1, pp. 99–112, Jan. 2020.
- [13] M. Zouina and B. Outtaj, “A novel lightweight URL phishing detection system using SVM and similarity index,” Hum.-Centric Comput. Inf. Sci., vol. 7, no. 1, p. 17, Jun. 2017.
- [14] R. Ø. Skotnes, “Management commitment and awareness creation—ICT safety and security in electric power supply network companies,” Inf. Comput. Secur., vol. 23, no. 3, pp. 302–316, Jul. 2015.
- [15] R. Prasad and V. Rohokale, “Cyber threats and attack overview,” in Cyber Security: The Lifeline of Information and Communication Technology. Cham, Switzerland: Springer, 2020, pp. 15–31.
- [16] T. Nathezhtha, D. Sangeetha, and V. Vaidehi, “WC-PAD: Web crawling based phishing attack



- detection,” in Proc. Int. Carnahan Conf. Secur. Technol. (ICCST), Oct. 2019, pp. 1–6.
- [17] R. Jenni and S. Shankar, “Review of various methods for phishing detection,” EAI Endorsed Trans. Energy Web, vol. 5, no. 20, Sep. 2018, Art. no. 155746.
- [18] (2020). Accessed: Jan. 2020. [Online]. Available: <https://catches-of-themonth-phishing-scams-for-january-2020>
- [19] S. Bell and P. Komisarczuk, “An analysis of phishing blacklists: Google safe browsing, OpenPhish, and PhishTank,” in Proc. Australas. Comput. Sci. Week Multiconf. (ACSW), Melbourne, VIC, Australia. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–11, Art. no. 3, doi: 10.1145/3373017.3373020.
- [20] A. K. Jain and B. Gupta, “PHISH-SAFE: URL features-based phishing detection system using machine learning,” in Cyber Security. Switzerland: Springer, 2018, pp. 467–474.
- [21] Y. Cao, W. Han, and Y. Le, “Anti-phishing based on automated individual white-list,” in Proc. 4th ACM Workshop Digit. Identity Manage., Oct. 2008, pp. 51–60.
- [22] G. Diksha and J. A. Kumar, “Mobile phishing attacks and defence mechanisms: State of art and open research challenges,” Comput. Secur., vol. 73, pp. 519–544, Mar. 2018.
- [23] M. Khonji, Y. Iraqi, and A. Jones, “Phishing detection: A literature survey,” IEEE Commun. Surveys Tuts., vol. 15, no. 4, pp. 2091–2121, 4th Quart, 2013.
- [24] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs, “Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions,” in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., Apr. 2010, pp. 373–382.
- [25] P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta, “PhishNet: Predictive blacklisting to detect phishing attacks,” in Proc. IEEE INFOCOM, Mar. 2010, pp. 1–5.
- [26] P. K. Sandhu and S. Singla, “Google safe browsing-web security,” in Proc. IJCSET, vol. 5, 2015, pp. 283–287.
- [27] M. Sharifi and S. H. Siadati, “A phishing sites blacklist generator,” in Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl., Mar. 2008, pp. 840–843.
- [28] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, “An empirical analysis of phishing blacklists,” in Proc. 6th Conf. Email Anti-Spam (CEAS), Mountain View, CA, USA. Pittsburgh, PA, USA: Carnegie Mellon Univ., Engineering and Public Policy, Jul. 2009.
- [29] 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.
- [30] G. Xiang, J. Hong, C. P. Rose, and L. Cranor, “CANTINA+: A featurerich machine learning



International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-**5.86**

framework for detecting phishing web sites,” ACM
Trans. Inf. Syst. Secur., vol. 14, no. 2, pp. 1–28, Sep.
2011