A Novel Approach For Phishing Detection System Using Hybrid Data Mining Techniques

Dr.G.Siva Nageswara Rao[1], J.Vijay Reddy[2],
[1]Professor, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. sivanags@kluniversity.in.
[2]PG student , Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.
vijayreddy8718@gmail.com

Abstract: *By using a large dataset based on the fishing url, they begin attacks on the Internet. The aim of the study is to improve detection of cyber hazards using different types of machine learning methods. These algorithms include "Decision Tree [4], Linear Regression [4], Random Forest [4], Naive Bayes, Gradient Boosting Classifier, Support Vector Classifier, and a new hybrid LSD model". We have used a hybrid model by combining the predictions of many individual models, such as a stacking classify, a ensemble technique. This model connects predictions with "Random Forest [4] Classifier and MLP Classifier as base classifiers". We have achieved it through carefully cross- fold validation and Grid Search Hyper parameter Optimization. As a meta-estimator, it appoints the LGBM classification to reach the final prediction, which extends the project's ability to perform better classification. The effect of the model is evaluated using matrix including F1 score, recalling, accuracy and accuracy. The results show that the Hybrid LSD model effectively reduces the risk of fish attacks and provides strong protection against the ever -changing cyber threats. This study contributes to the development of better cyber security measures, and shows how you can improve the safety of the Internet by learning machine.*

"Index Terms: *Phishing attacks, Machine learning algorithms, Cyber threat detection, Hybrid LSD model, Cyber security measures".*

INTRODUCTION

A smart danger on web is fishing, where thieves stated as legitimate businesses or websites in an attempt towards obtain important information (such as password, credit card details or personal history). In order towards avoid financial losses & towards ensure that sensitive information does not get into wrong hands, it is extremely important towards detect fishing efforts. fight against fishing is equipped among machine learning, a kind of help. It detects fish efforts by analyzing data from large & scale, finding patterns in it & using this knowledge. An important advantage is that ML systems abide extremely flexible, as they can meet new & changing fishing efforts. Checking URL, or site address, is a technique for identifying fishing efforts. incorrect or misspelled domain names or excessive number of under domain enemies abide common url tabs. Such a nice irregularity is quite easy towards detect machine learning algorithm. successful fish declaration system can easily endure combined among a wide range of web-based applications, including email clients, corporate networks & browser. These interconnected systems abide always looking for new fishing efforts & protecting users immediately against them. Internet has evolved into an integrated component of modern life, thanks towards progress of communication & information technology. It facilitates surplus of life improvement opportunities in communication, entertainment, education, retail medicine. Criminals see Internet as a way towards take their physical crimes online because our online life develops. Although there abide many positive benefits towards using Internet, some abide also negative, such as oblivion it gives users. According towards research from Partsmouth University (2016), Raguchi & Robila (2006), & Hong (2012), individuals & organizations lose millions of dollars every day. Cybercrime, one of most basic forms of fishing, grows at an exponential speed. [12] Only players have spread among expansion of Internet time. among extensive use of Internet, fishing attacks have increased in popularity. One of main methods is utilized by playing weaknesses. People who suffer from fishing fraud sometimes come for fraud as websites used towards fool them or look like other popular sites. For most parts, unskilled internet users cannot show difference between legitimate & dangerous websites. Because of this, fishing blacklists were developed. Fishing blacklists abide databases among malicious software maintained by experts. They enable ordinary people towards learn about fishing sites that they could travel [18]

LITERATURE SURVEY

Introduction towards "Phishpedia", a groundbreaking logo-based phishing identity system that accounts for its remarkable accuracy & minimal impact on driving time. author of system, Y. Lynn, R. Liu, D. M. Diwakaran, J. Y. Ng, Q. Z Chan, Y. Lu, Y. C., F. Zhang & JS abide Dongs. Compared towards existing methods, our condition achieves -Art -Pype learning system better results in identifying phishing efforts correctly, especially when it comes towards recognition & matching. Not only does it work better than current methods, but it also finds fishing spots that were not before, which makes Armed Forces much stronger against fishing efforts. When it comes towards improving cyber security, Phishpedia is in its own league. Negative: effectiveness of Phishpedia depends on presence & quality of people on websites. Sometimes there is regular upgrading, & maintenance must endure ahead of changed phishing strategies. [1]

By using artificial nerve networks (Ann) towards analyze HTML & URL properties, introduce a groundbreaking algorithm for Shirazi, Hens & Raya Mobile Phishpedia Detection. Modern deep transformers such as Burt, Electra, Robert & Mobile Bert have been included in their method of effective learning from URL text. state -Art -art system effectively administers fast training, smooth maintenance & real -time delaying on mobile devices. It guarantees top -oriented performance, strengthens prevention against fish attacks & maximizes use of resources for better mobile cyber security. Negative: Complicated fishing on actual pages cannot go towards anyone's attention if URL is only way towards detect. availability & quality of pre -trained transformers may vary. [2] A. Akanka's dissertation examines SSL certificates used by fishing spots, analyzing properties of attackers & developing an auto-detection system that uses these aspects. Research introduces a groundbreaking SSL certificate- based fish-declaration system that uses decision tree [4] machine learning due towards its openness & efficiency. system claims excellent accuracy & has a user-friendly web api. This letter presents a holistic approach towards cyber security problems & highlights need for future adjustment towards develop a fishing strategy & guarantee continuous system updates. Lack of Missing: If malicious actors find ways towards copy real SSL certificates, efficiency of system can endure compromised. scalability of system towards handle many domains is barely affected. Third Logistics regression, decision tree [4], neural networks & Gaussian naive Bayes, H. Shahiriyar & S. Using machine learning methods that Nimgadda focuses on network infiltration system (IDS) in their joint efforts. Finding specific & unusual network behavior, especially TCP/IP teams, is primary goal of research. Decision Tree [4] works on a publicly available dataset, but authors emphasized need for evaluation & real world testing towards confirm its efficiency & accuracy in Real -world network infiltration scenarios. Lack of deficiencies: results may not endure reflective towards real world or changing dangers. This algorithm choice is not overall; Alternative approaches can produce different results. [4]

In his groundbreaking work, A. of. Dutta creates a refined system for detecting fishing sites using a monitored machine learning technique random forest [4]. This process examines & selects relevant features that carefully separate fishing sites. solution, when distributed as a smart browser plugin, detects phishing sites among 98.8 percent accuracy, towards handle human deficiencies in online security, continuously. main goal is towards improve Internet security & give users a strong security against potential cyber dangers, even though it causes a false alarm sometimes. Resistance: quality of compatibility plants for new phishing strategies is influenced by quality of facilities. If they lie, users may lose confidence in results [5]

METHODOLOGY

Proposed work

The proposed technique uses a state-of-the-art hybrid machine learning method towards identify fish attacks using URL properties. This strengthens rescue against attacks & defends users by using a wide range of machine learning techniques. Grid search hyper parameter optimization & cross-fold validation, when common, prediction accuracy is greatly improved. expansion of project adds a stacking classifies towards create a hybrid model, which further improves possibilities. In this artist's contingent, two base classifies - Random Forest [4] Classifier & MLP Classifier— -to create a strong general model. classification performance of project is improved by incorporating LGBM classification as a meta-estimator, which refines final prediction. This all -encompassing method represents an important step in cyber security by guaranteeing an effective & reliable defense system against fish attacks.

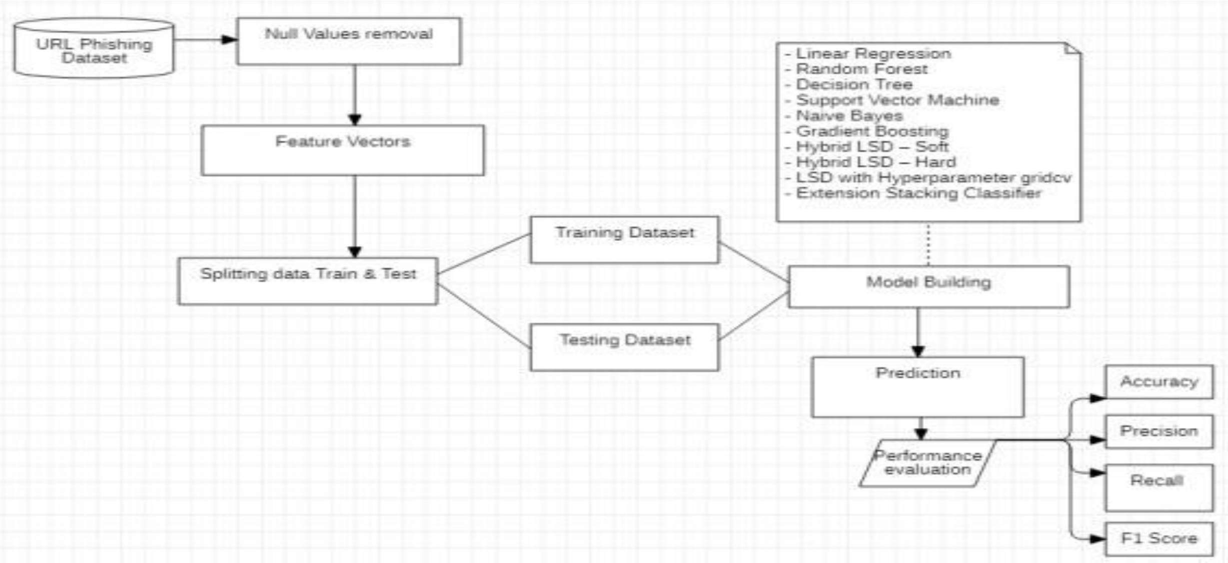
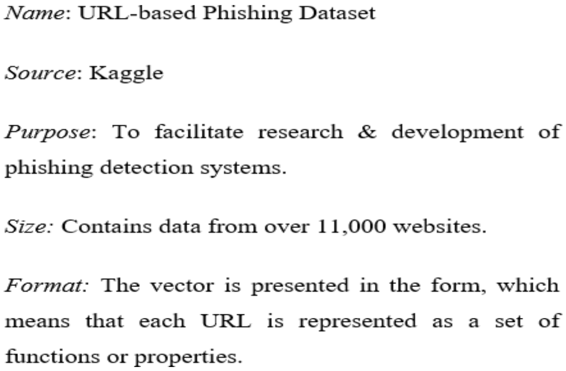


Fig 1: System Architecture

A) Dataset Collection

Research & system for development purposes, "URL-based fish dataset" collects information on malicious & valid URL. It comes from Kaggle, a well-known market for data set & computer science competitions. dataset is described in wide stroke here:



Name: URL-based Phishing Dataset

Source: Kaggle

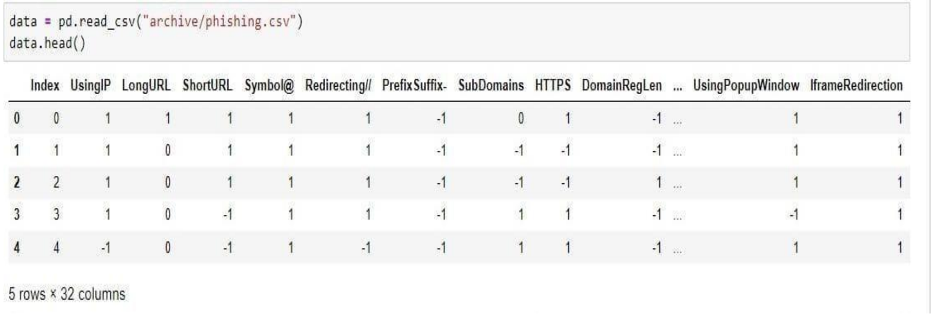
Purpose: To facilitate research & development of phishing detection systems.

Size: Contains data from over 11,000 websites.

Format: The vector is presented in the form, which means that each URL is represented as a set of functions or properties.

Machine learning models can use functions associated among each URL (in vector form), towards determine if a certain URL is related towards fishing; This information is likely towards endure available in dataset, which is structured as each entry or example matches a URL.

Some examples of specific parameters involved in fish dataset for fish duty include domain age, https use, keyword density, url length & presence or absence of specific keywords. Machine learning models rely on these qualities towards learn towards distinguish between real & malicious URLs.



```
data = pd.read_csv("archive/phishing.csv")
data.head()
```

	Index	UsingIP	LongURL	ShortURL	Symbol@	Redirecting//	PrefixSuffix-	SubDomains	HTTPS	DomainRegLen	...	UsingPopupWindow	IframeRedirection
0	0	1	1	1	1	1	-1	0	1	-1	...	1	1
1	1	1	0	1	1	1	-1	-1	-1	-1	...	1	1
2	2	1	0	1	1	1	-1	-1	-1	1	...	1	1
3	3	1	0	-1	1	1	-1	1	1	-1	...	-1	1
4	4	-1	0	-1	1	-1	-1	1	1	-1	...	1	1

5 rows × 32 columns

B) Pre-processing

Using Panda's data frames: towards clean, replace & prepare dataset at this stage, we use panda, a strong python data manipulation library. This includes addressing missing values, changing data formats & organizing information for further modeling or analysis.

Visualization among seaborn & food matplotlib: towards learn more about properties of dataset, we use seaborn & food matplotlib towards create visualizations such as diagrams & graphs. In order towards make well - informed decisions for further analysis, this phase helps us understand distribution, relationships & patterns present in data.

Labeling: towards translate classified label into numerical values, we use a label codance, a pre -rosaring method. Given that machine learning model usually requires numerical entrance, it is necessary. dataset is guaranteed by possibility of model for understanding & learning models from classified data.

Feature selection: most relevant properties from dataset abide found & selected at this stage. By focusing most useful variable & reducing noise, construction choices abide needed towards increase model performance. model can endure found using methods such as statistical testing, correlation analysis or machine learning techniques.

C) Training & Testing

Our first machine learning model, Model 9, was used on preprisable data sets during initial phase of project for analysis & interpretation. After that, we wanted towards improve accuracy of our predictions during extension phase, so we created a hybrid model, which combined output from other models. By combining best features of many models, this new strategy expects towards increase accuracy of our predictions. At same time, we created an authentication dream, flask -based fronts, so that users interact among model. A user -friendly & available interface is provided by this friend, so users can enter & easily rebuild predictions. towards train first indicated machine learning models towards detect complex computer patterns & correlations, we will use Predicable Dataset. This will endure basis for our project. A separate test is fully evaluated on a separate test data set after training process. towards evaluate effect of these models in identifying Phishing urls, performance matrix is used as recalling, accuracy & F1 score carefully. This perfect assessment process acts as an essential quality control measure, guarantees that models show purity & addiction, & confirm their suitability for practical use. aim of our project is towards provide reliable & advanced phishing url identity using this all-encompassing methodology.

D) Algorithms.

Stacking Classifier: As a fundamental classifies, a stacking classifies from Project Random Forest combines [4] Classifier & MLP Classifier predictions using a cloth approach. It extends project functions for better classification performance using LGBM classifies as a meta-estimator towards make final prediction.

LSD: Hyperparameterized logistic regression, supporting voter & decision -making wood [4] Model GridcV is a hybrid classification model that improves accurately & efficiency by combining benefits of decision tree [4], support vector machine & logistical regional methods. GRIDCV is useful for different types of classification problems because it looks methodical through hyperparameter combinations towards maximize model performance.

Hybrid LSD (HARD): towards make classification decisions, use Hybrid LSD (Hard) Model Decision Tree [4] A hard voice method in combination among algorithm of Support Vector Machine & Logistic Region. In order towards improve accuracy & strength of different types of classification functions, each component model provides a prediction, & final alternative is set by a majority vote.

Hybrid LSD (Soft): This model uses soft voting towards classify data by mixing decision tree, supporting vector machine & logistic region [4]. It improves accuracy of classification job by using possibilities of each model towards generate prophecies while maintaining adaptation towards handle a variety of inputs.

Gradient Boosting: This outfit machinery method combines skills of many weak students, who usually determine trees, gradually towards build a prediction model [4]. It completes it by focusing on errors generated by previous models & changing predictions towards reduce these errors, eventually provides a strong & accurate future model that works very well in different types of functions, such as regression & classification.

Random forest: Random forest [4] is a dress learning technique that generates predictions by adding many decision trees [4]. This decision is driven by average of predictions from a group of trees [4] trained on arbitrary selection of data. For both classification & regression problems, this dress technique provides strong performance, reducing overfitting & improving accuracy.

To classify or predict results, a machine learning model called Decision Tree [4] shares most important feature shares input towards earliest. It produces a three - -like structure that is useful & explanatory for a variety of functions, each branch represents a possible option & each node represents a feature.

Support Vector Classifier: A Support Vector Classifier (SVC) is a machine learning model that maximizes margins between many data sections by determining optimal limit (hyperplane). This works well for both binary & multi - class classification applications, as it recognizes important support vector towards produce accurate classification.

A classification process called logistic regression estimates possibility that an entrance will fall into a particular category. towards classify entrance towards one of two or more categories, a threshold is used after Sigmoid feature, & maps input functions for a probability value between 0 & 1. towards do data best & produce an accurate classification, coefficient model learns coefficient during training.

The convenience implements possible classification technique known as Naïve bayes, using "naive" freedom base. Depending on possibilities of each component properties, it determines possibility that a data point belongs towards a specific class. Lesson classification, spam detection & other landscapes where freedom of convenience is a commendable contact, Bole Bayes is especially effective.

EXPERIMENTAL RESULTS

Accuracy Val: A test ability towards make a proper difference between healthy & sick cases is a measure of accuracy. We can determine accuracy of a test through calculating proportion of cases undergoing proper positivity & genuine negative. It is possible towards express this mathematically:

$$Accuracy\ Val = \frac{TP1 + TN1}{TP1 + FP1 + TN1 + FN1} \qquad (1)$$

Precision Val: relationship between events or tests certain abide properly classified towards anyone classified as positive is called accurate. Therefore, there is a formula considering determining accuracy:

$$Precision\ val = \frac{TP1}{TP1 + FP1} \qquad (2)$$

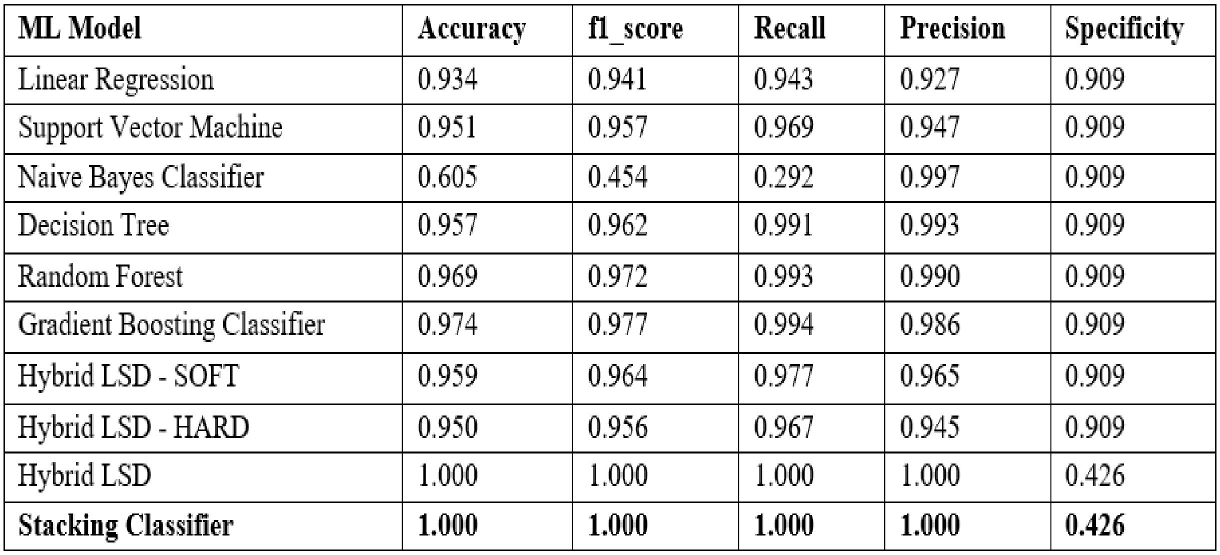
Recall Val: In machine learning, recall is a solution towards how well a model can find all examples of a specific class. ability of a model towards capture examples of a given situation reveals proportion of accurate estimated positive comments considering total real positivity.

$$Recall\ Val = \frac{TP1}{TP1 + FN1} \quad (3)$$

F1-Score Val: F1 score is a measure towards evaluate purity of a model in machine learning. It takes memory & accuracy of a model & mixes them. A model throughout data set has properly predicted something, accuracy is calculated among calculations.

$$F1\ Score = 2 * \frac{Recall\ Val\ X\ Precision\ Val}{Recall\ Val + Precision\ Val} * 100 \quad (1)$$

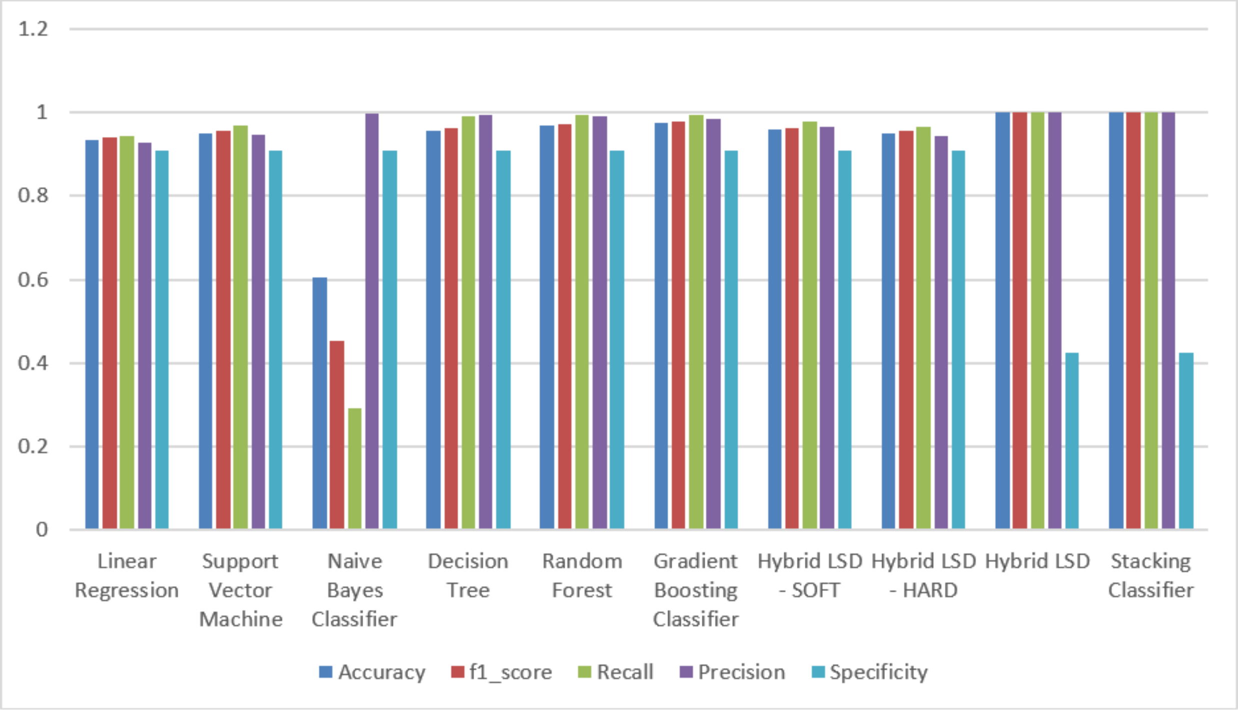
Table 1 assesses performance parameters of each algorithm, including accuracy, accurate, recall and F1 score. Stacking classifier is constantly performing better than all other algorithms in all matrix. In addition, calculations are compared towards different methods in tables.



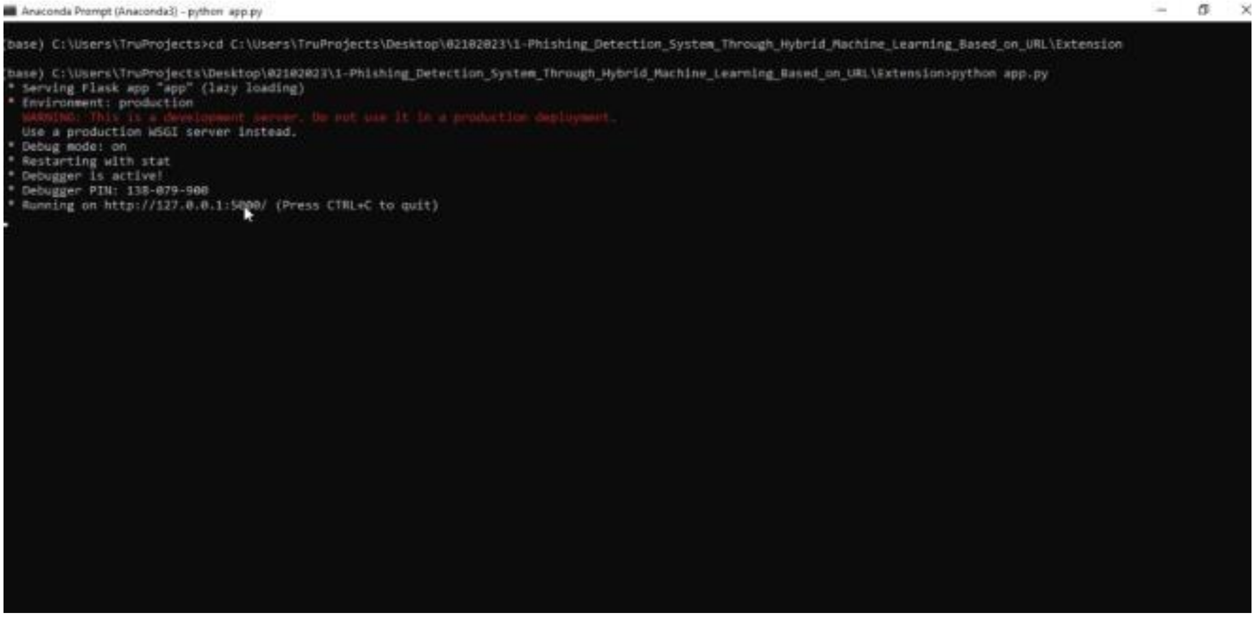
ML Model	Accuracy	f1_score	Recall	Precision	Specificity
Linear Regression	0.934	0.941	0.943	0.927	0.909
Support Vector Machine	0.951	0.957	0.969	0.947	0.909
Naive Bayes Classifier	0.605	0.454	0.292	0.997	0.909
Decision Tree	0.957	0.962	0.991	0.993	0.909
Random Forest	0.969	0.972	0.993	0.990	0.909
Gradient Boosting Classifier	0.974	0.977	0.994	0.986	0.909
Hybrid LSD - SOFT	0.959	0.964	0.977	0.965	0.909
Hybrid LSD - HARD	0.950	0.956	0.967	0.945	0.909
Hybrid LSD	1.000	1.000	1.000	1.000	0.426
Stacking Classifier	**1.000**	**1.000**	**1.000**	**1.000**	**0.426**

Table.1 Performance Evaluation Table

Graph.1 Comparison Graph

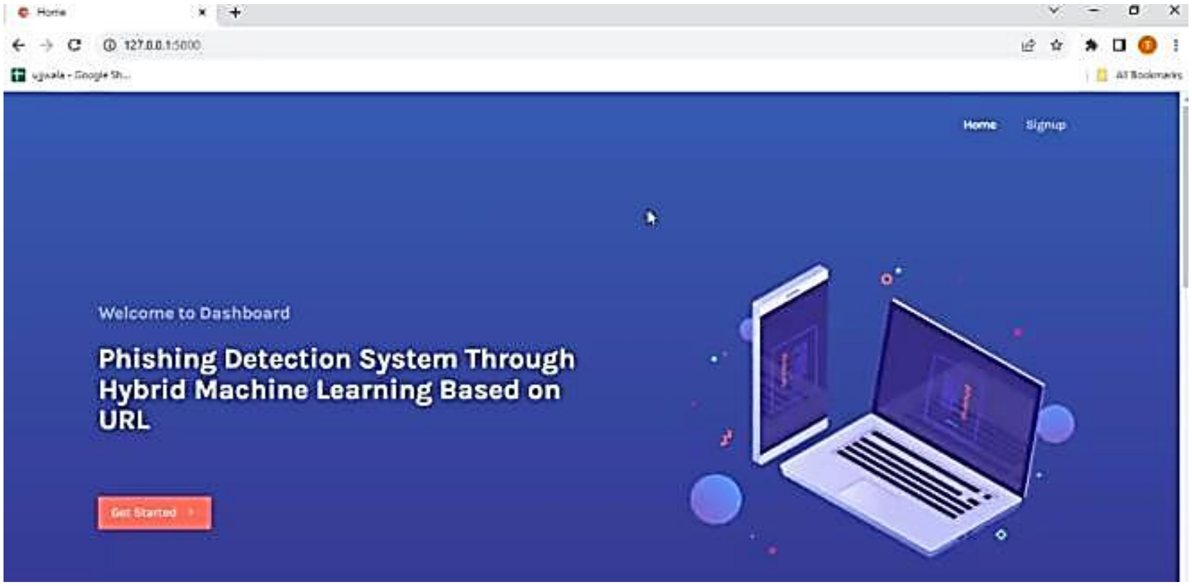


Graph (1) shows specificity sky blue, accuracy in blue, F1-Score in red, recall in green, & precision in purple. In every metric, Stacking Classifier performs better than other models, attaining highest scores. These findings abide graphically depicted in graphs above.



```
(base) C:\Users\TruProjects>cd C:\Users\TruProjects\Desktop\02102023\1-Phishing_Detection_System_Through_Hybrid_Machine_Learning_Based_on_URL\Extension

(base) C:\Users\TruProjects\Desktop\02102023\1-Phishing_Detection_System_Through_Hybrid_Machine_Learning_Based_on_URL\Extension>python app.py
 * Serving Flask app "app" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: on
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: 138-079-900
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

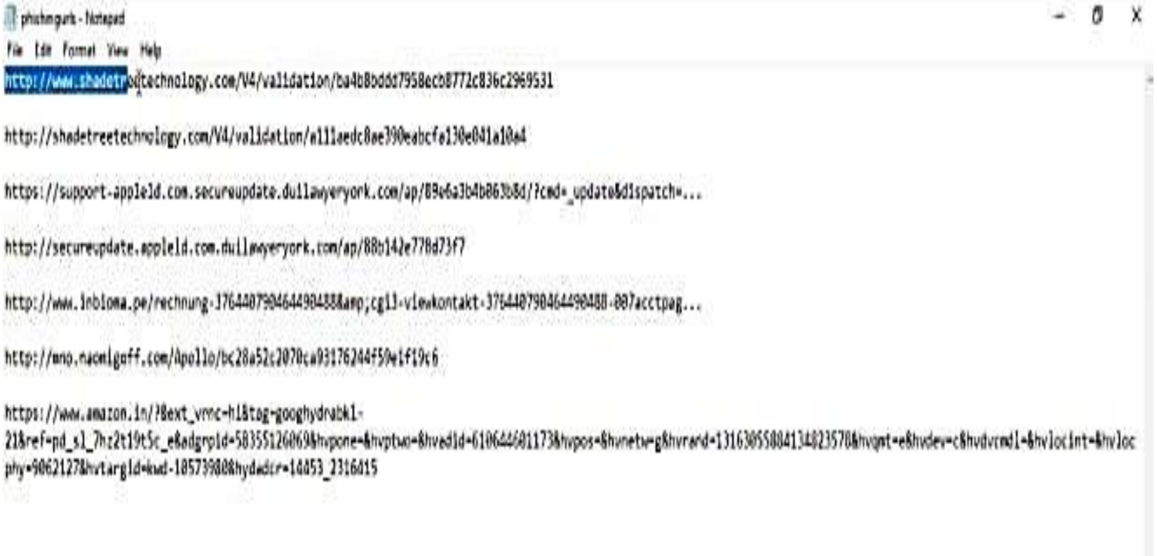
Fig 1: URL Link towards Web Page



Welcome to Dashboard

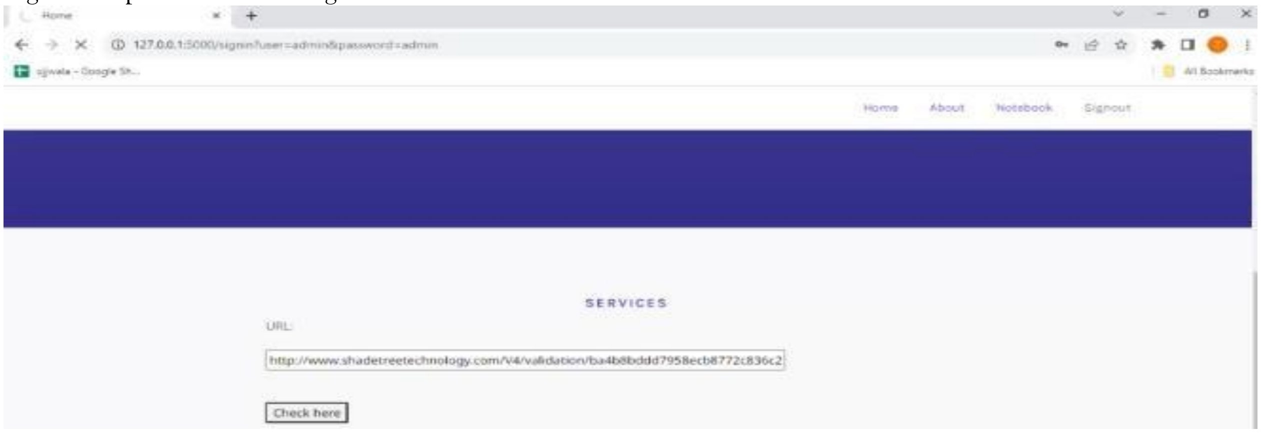
Phishing Detection System Through Hybrid Machine Learning Based on URL

Fig 2: Home page



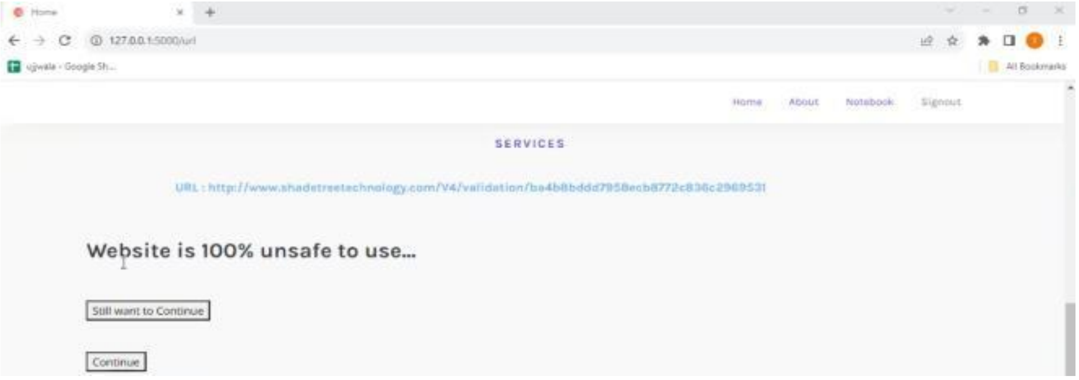
```
http://www.shadetreetechnology.com/V4/validation/ba4b8bddd7958ecb8772c836c2969531

http://shadetreetechnology.com/V4/validation/a111aedc8ae390eabcfa130e041a10a4

https://support-appleld.com.secureupdate.duilawyeryork.com/ap/89e6a3b4b063b8d/?cmd=_update&dispatch=...

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

http://www.inbioma.pe/rechnung-376440790464490488&amp;cgi1-viewkontakt-376440790464490488-007acctpag...

http://mno.naomigoff.com/Apollo/bc28a52c2070ca93176244f59e1f19c6

https://www.amazon.in/?&ext_vrnc=hi&tag=googhydrabk1-
21&ref=pd_sl_7hz2t19t5c_e&adgrpid=58355126069&hvpone=&hvptwo=&hvadid=610644601173&hvpos=&hvnetw=g&hvrand=13163055804134823578&hvqmt=e&hvdev=c&hvdvcmdl=&hvlocint=&hvloc
phy=9062127&hvtargid=kwd-10573980&hydadcr=14453_2316415
```

Fig 3: Sample data for testing



URL: http://www.shadetreetechnology.com/V4/validation/ba4b8bddd7958ecb8772c836c2

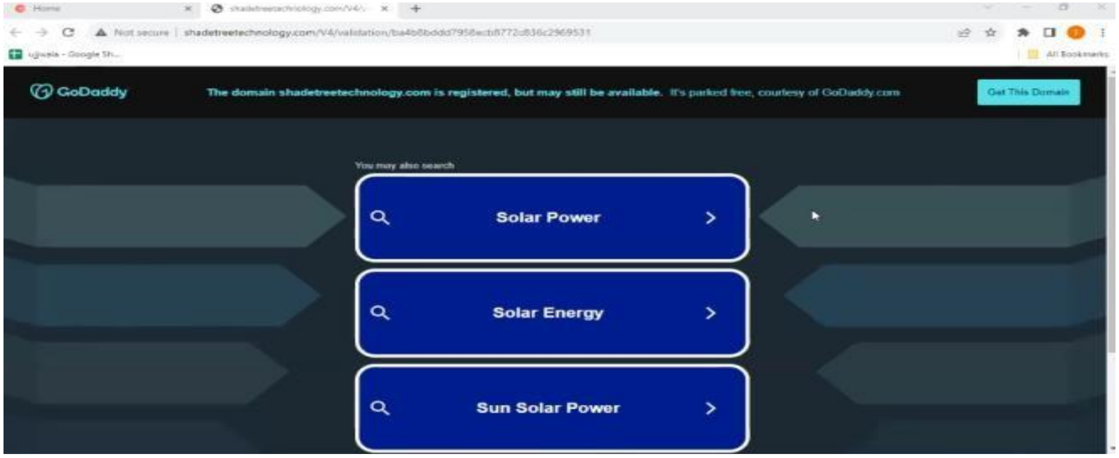
Fig 4: Entered URL



URL : http://www.shadetreetechnology.com/V4/validation/ba4b8bddd7958ecb8772c836c2969531

Website is 100% unsafe to use...

Fig 5: URL result unsafe 100%



The domain shadetreetechnology.com is registered, but may still be available. It's parked free, courtesy of GoDaddy.com

Fig 6: Search Other Urls too

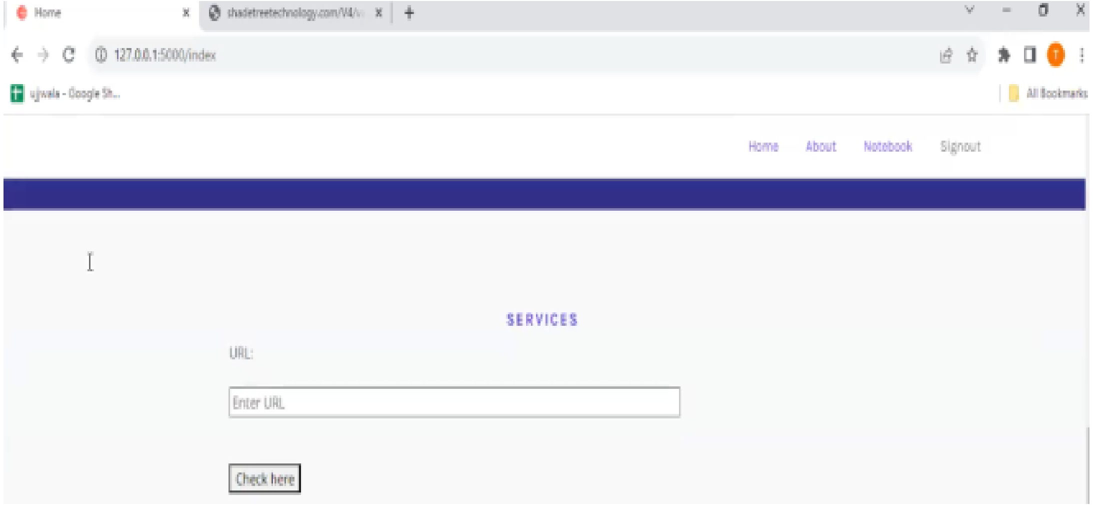
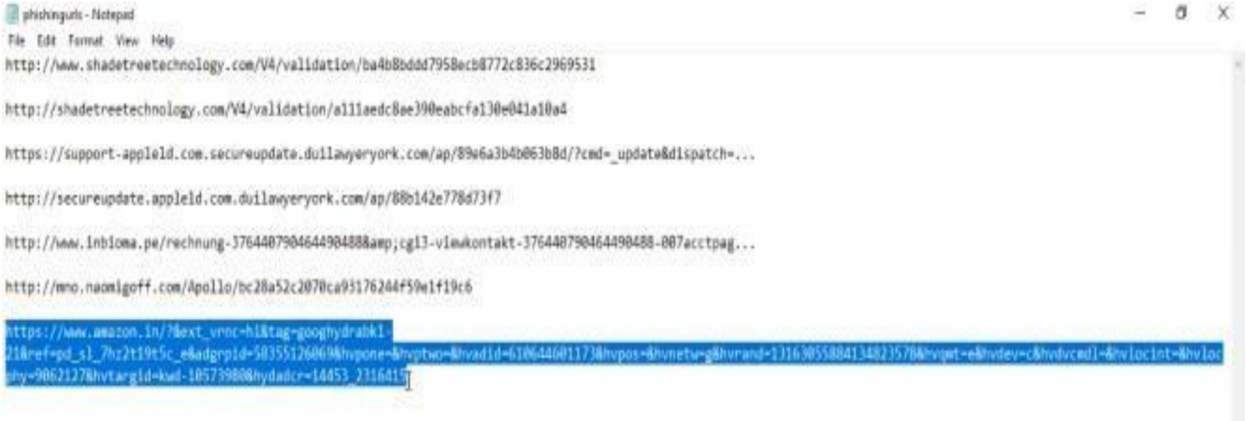
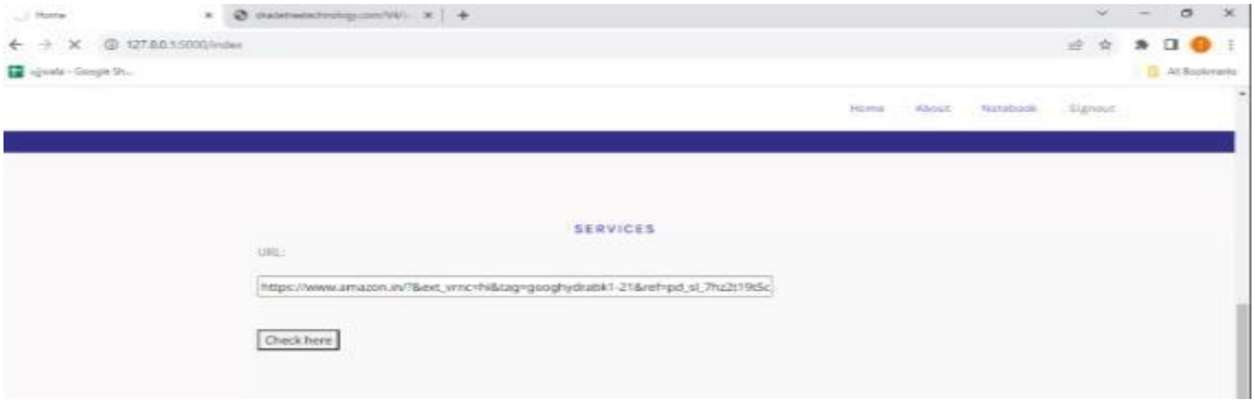


Fig 7: Enter New URL



```
http://www.shadetreetechnology.com/V4/validation/ba4b8bddd7958ecb8772c836c2969531

http://shadetreetechnology.com/V4/validation/a111aedc8ae390eabcfa130e041a10a4

https://support-appleld.com.secureupdate.duilawyeryork.com/ap/89e6a3b4b063b8d/?cmd=_update&dispatch=...

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

http://www.inbioma.pe/rechnung-37644079046449048&amp;cgi3-viewkontakt-376440790464490488-007acctpag...

http://mno.naomigoff.com/Apollo/bc28a52c2070ca93176244f59e1f19c6

https://www.amazon.in/?&ext_vrnc=hi&tag=googhydrabk1-
21&ref=pd_sl_7hz2t19t5c_e&adgrpid=58355126069&hvpone=&hvptwo=&hvadid=610644601173&hvpos=&hvnetw=g&hvrand=13163055884134823578&hvqmt=e&hvdev=c&hvdvcmdl=&hvlocint=&hvloc
phy=9062127&hvtargid=kwd-10573980&hydadcr=14453_2316415
```

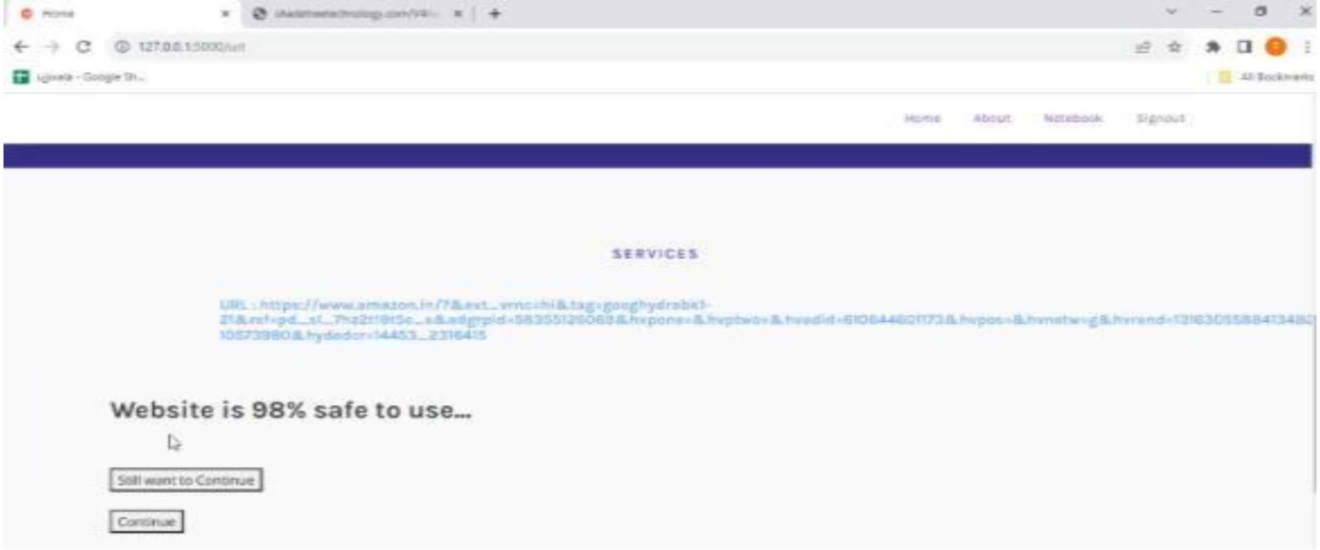
Fig 8: Sample data for testing



URL:

https://www.amazon.in/?&ext_vrnc=hi&tag=googhydrabk1-21&ref=pd_sl_7hz2t19t5c

Fig 9: Entered New URL



URL : https://www.amazon.in/?&ext_vrnc=hi&tag=googhydrabk1-
21&ref=pd_sl_7hz2t19t5c_e&adgrpid=58355126069&hvpone=&hvptwo=&hvadid=610644601173&hvpos=&hvnetw=g&hvrand=1316305588413482
10573980&hydadcr=14453_2316415

Website is 98% safe to use...

Fig 10: URL result page (safe/ unsafe)

CONCLUSION

With help of a hybrid machine learning strategy, project was able towards prefer URL properties for fish decisions, & it was a success. significant benefit in accuracy & efficiency was obtained by system through use of different models, including "Decision Tree [4] s, Random Forest [4] s, support vector classifiers, & an LSD-based stacking classifier". By using an expansion stacking classifies, fish declaration system was greatly improved, which was responsible for its high accuracy & F-point. This all-dedicated method provides a strong defense against refined fishing efforts, & solves a big question in cyber security. A better degree of adaptability towards develop fishing techniques was ensured by integrating different machine learning models, which brought diversity towards system's abilities. As a result of its positive effects on accuracy & efficiency, project has opportunity towards strengthen cyber security measures & make a significant contribution in fight against cyber threats. among growing processing of fish attacks, proposed system proves a strong defense mechanism, its practical use shows towards protect sensitive data & reduce cyber security risk.

FUTURE SCOPE

Constant improvement & adjustments in new fishing techniques abide in future plans for this project. In order towards improve active safety opportunities of system, future studies can use deep learning, behavioral analysis & real -time panties intelligence. Working among professionals for cyber security & other industry players can also help create a strong solution. Access towards this system can endure expanded by checking blame environment & distribution on Internet of Things devices, as well as by developing a user -friendly interface. A top modern solution in ever -existing scope of cyber security, model undergoes constant changes towards change hazard landscape, guarantees its constant effect.

REFERENCES

[1] Y. Lin, R. Liu, D. M. Divakaran, J. Y. Ng, Q. Z. Chan, Y. Lu, Y. Si, F. Zhang, & J. S. Dong, "Phishpedia: A hybrid deep learning based approach towards visually identify phishing webpages," in Proc. 30th USENIX Secur. Symp. (USENIX Security), 2021, pp. 3793–3810.

[2] H. Shirazia, K. Haynesb, & 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 & S. Nimmagadda, "Network intrusion detection for TCP/IP packets among machine learning techniques," in Machine Intelligence & 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 & 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, & M. Supramaniam, "Phishing website detection: An improved accuracy through feature selection & ensemble learning," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 1, pp. 252– 257, 2019.

[8] A. Aggarwal, A. Rajadesingan, & 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, & E. Snekkenes, Computer Security— ESORICS 2017, vol. 10492. Oslo, Norway: Springer, Sep. 2017.

[10] P. George & P. Vinod, "Composite email features for spam identification," in Cyber Security. Singapore: Springer, 2018, pp. 281–289.

[11] H. S. Hota, A. K. Shrivas, & R. Hota, "An ensemble model for detecting phishing attack among proposed remove-replace feature selection technique," Proc. Comput. Sci., vol. 132, pp. 900–907, Jan. 2018.

[12] G. Sonowal & K. S. Kuppusamy, "PhiDMA—A phishing detection model among multi-filter approach," J. King Saud Univ., Comput. Inf. Sci., vol. 32, no. 1, pp. 99–112, Jan. 2020.

[13] M. Zouina & B. Outtaj, "A novel lightweight URL phishing detection system using SVM & similarity index," Hum.-Centric Comput. Inf. Sci., vol. 7, no. 1, p. 17, Jun. 2017.

[14] R. Ø. Skotnes, "Management commitment & awareness creation—ICT safety & security in electric power supply network companies," Inf. Comput. Secur., vol. 23, no. 3, pp. 302–316, Jul. 2015.

[15] R. Prasad & V. Rohokale, "Cyber threats & attack overview," in Cyber Security: Lifeline of Information & Communication Technology. Cham, Switzerland: Springer, 2020, pp. 15–31.

[16] T. Nathezhtha, D. Sangeetha, & 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 & 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 & P. Komisarczuk, "An analysis of phishing blacklists: Google safe browsing, OpenPhish, & 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 & 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, & 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 & J. A. Kumar, "Mobile phishing attacks & defence mechanisms: State of art & open research challenges," Comput. Secur., vol. 73, pp. 519–544, Mar. 2018.

[23] M. Khonji, Y. Iraqi, & 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, & J. Downs, "Who falls for phish? A demographic analysis of phishing susceptibility & effectiveness of interventions," in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., Apr. 2010, pp. 373–382.

[25] P. Prakash, M. Kumar, R. R. Kompella, & M. Gupta, "PhishNet: Predictive blacklisting towards detect phishing attacks," in Proc. IEEE INFOCOM, Mar. 2010, pp. 1–5.

[26] P. K. Sandhu & S. Singla, "Google safe browsing-web security," in Proc. IJCSET, vol. 5, 2015, pp. 283–287.

[27] M. Sharifi & 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, & 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 & Public Policy, Jul. 2009.

[29] Y. Zhang, J. I. Hong, & L. F. Cranor, "Cantina: A content-based approach towards 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, & L. Cranor, "CANTINA+: A featurerich machine learning framework for detecting phishing web sites," ACM Trans. Inf. Syst. Secur., vol. 14, no. 2, pp. 1–28, Sep. 2011