ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 3, 2025

PHISHING DETECTION SYSTEM THROUGH HYBRID MACHINE LEARNING BASED ON URL

[1,] Dr. N. SRIVANI (Associate Professor) [2,] VANGAPALLY MANI SAI VARMA

CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA-MEDCHAL

(COMPUTER SCIENCE AND ENGINEERING)

Abstract - Phishing attempts on the internet using a whole dataset of phishing URLs. The study's goal is to improve cyber threat detection by using "a variety of ML methods, such as decision Tree [4], Linear Regression[4], Random forest [4], Naive Bayes, Gradient Boosting Classifier, support Vector Classifier, and a new hybrid LSD model." We used careful cross validation and optimization of the grid search hyperparameters. We used a hybrid model that combined predictions of several individual models, "such as stacking classifier, file technique that combines predictions from random forest classifier [4] and MLP classifier as basic classifiers." To create a final prediction, the LGBM classifier uses a meta-testimator. This contributes to the project functions and improves categorization better. "We use rating metrics, such as precision, accuracy, recall, and F1-score to see how well the model works." The results show that the LSD hybrid model is very good to stop phishing threats and is a strong obstacle to new cyber threats. This study helps make cybersecurity better and shows how ML could help make the internet safer.

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

I. INTRODUCTION

Phishing is a clever online threat that hackers designate as being a trusted source, bank or website to provide "personal information such as passwords, credit card numbers, and personal information." It is important to decide your phishing effort, as important information is retained from being incorrectly obtained and protecting yourself from money. A kind of artificial intelligence ML is very good at stopping phishing. We look through a lot of data, "learn patterns from it, and use this information to find phishing attempts. The big advantage is that ML systems learn new phishing methods" and adapt to them, which makes them very strong. The approach to finding phishing is to look at the address or URL of your website. Phisher Musik often brings URLs by using incorrect calculation domain names or adding too many subdums. The ML model is pretty good to find these small problems. You can add a set of online tools, including web browsers, email customers, and corporate networks without any problems. These technologies work together in real time, constantly looking for phishing threats to incoming data and protecting users.

The internet has become a necessary component of our life in this age of technology. It makes communication, entertainment, education, shopping, and other things much easier in our life. As we move more of our lives online, thieves see the internet as a way to move their real-world crimes into a digital

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

space. The internet is useful in many ways, but it also has some problems, like the fact that it lets people stay anonymous. [7] as the number of people who use the internet grows quickly, so is the number of cybercrimes. "Every day, people and groups lose millions of dollars (Hong, 2012; Ragucci and Robila, 2006; university of Portsmouth, 2016). Phishing is one of the most common types of cybercrime, and it is getting worse every day. [12]" as the digital age has grown, so have the number of bad people. in the time when websites are a part of everyday life, phishing assaults become popular. Taking advantage of people's flaws is a big reason why consumers are victimized. Phishing websites are similar to real websites and other websites, which makes many people fall in love with jokes. Bad websites can be similar to good, so those who are professionals who use the internet can't recognize the difference. This formed a phishing blacklist. Experts maintain a phishing blacklist, software data. They help ordinary people learn about phishing websites that they may be able to visit. [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 introduce "Phishpedia," a groundbreaking logo-based phishing identification system that is very accurate and has no effect on runtime." This new DL algorithm is better than current methods at accurately identifying phishing attempts, especially when it comes to recognizing and matching logos. Its ability not only beats other methods, but it also finds phishing sites that have never been found before, making defenses against phishing attempts stronger. "Phishpedia is a one-of-a-kind and powerful tool for improving cybersecurity. Cons: Phishpedia's achievement depends on the quality and availability of logos on websites. To keep up with changing phishing methods," you need to keep your software up to date and maintained. [1]

"Shirazi, Haynes, and Raya have created a groundbreaking mobile-friendly phishing detection method that uses artificial Neural Networks (ANNs) to look at URL and HTML properties. Their method uses the latest deep transformers, such BERT, ELECTRA, RoBERTa, and MobileBERT, to learn" from URL content in a manner that is quick and easy. the new solution makes training quick, maintenance easy, and deployment on mobile devices in real time, which solves mobile security problems very well. This makes sure that the system works well in competition, sets up a strong defense against phishing threats, and makes the best use of resources to make mobile platforms more secure. Cons: it can only find URLs, therefore it could miss complicated phishing on real pages. It depends on pre-trained transformers, which can vary in quality and availability. [2]

A. Akanchha's thesis looks into SSL certificates on phishing sites, looking at the traits of attackers and creating an auto-detection system based on the properties of SSL certificates. The research uses decision Tree [4] ML since it is clear and works well. It introduces a new SSL certificate-based phishing detection system that is quite accurate and has an easy-to-use web API. The work shows how important it is to make changes in the future to keep up with new phishing methods and make sure that systems are always up to date. this is a full approach to dealing with cybersecurity problems. Cons: The system only works well provided the SSL certificate properties are correct. provided hackers find new ways to fake real certificates, this could be a problem.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

There isn't a lot of talk about how well the system can handle a lot of domains. [3]

"H. Shahriar and S. Nimmagadda worked together on a chapter that focuses on network Intrusion Detection systems (IDS) that use ML methods including Gaussian Naive Bayes, logistic regression, decision Tree [4], and neural networks." The goal of the study is to find out what regular and strange network behavior are, especially between TCP/IP layers. The decision -making tree [4] is doing quite well on public data files, but the authors emphasize that it is necessary to test in the real world and check the scalability in order to fully demonstrate its accuracy and efficiency in real detection of network. Disadvantages: The evaluation may not show how things really and how attacks are changing. Not all algorithms are given; Different approaches can bring different results. [4]

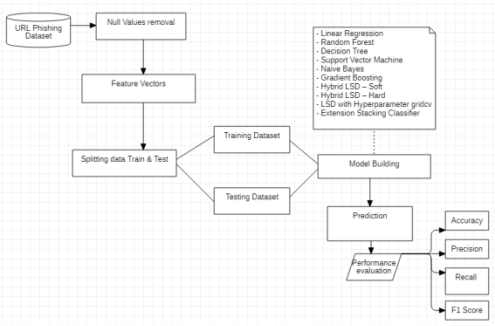
"A. k. Dutta's new method uses Random forest [4], a supervised ML methodology," to build a more powerful system that finds phishing websites. The strategy requires careful study and choice of relevant aspects that clearly set phishing sites apart. "The method is built into a smart browser extension and has an astounding 98.8% accuracy rate for finding phishing sites." this is a smart way to deal with people's weaknesses in online security. the main purpose is to greatly improve online security measures and give users a strong protection against possible cyber dangers, even though they may sometimes send out false signals. disadvantages: The quality of a feature affects how easily it can adapt to new phishing methods. The possibility of erroneous findings makes users less likely to trust the system. [5]

III. METHODOLOGY

"Modules:"

- Getting the packages You need
- "looking at the dataset—Phishing URL feature data"
- "Processing the data using Pandas data frame"
- "Making graphs with seaborn and matplotlib"
- "Encoding labels with Label Encoder"
- choosing features
- "Splitting the data into training and testing sets"
- training and building the model
- using the trained model to make predictions
- showing the final result through the front-end

A) "System Architecture"



"Fig 1: System Architecture"

"Proposed work"

"The proposed system uses a new hybrid ML method to find phishing attacks based on URL attributes. Various" ML methods provide increased protection against threats and protect users. By using a cross-compartment validation and a lattice search

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

hyperparameter optimization approach, predictions are made much more accurate. “This project will receive a stacking classifier that will do more, which is called a hybrid model. This ensemble method uses Random Forest [] classifiers and MLP classifiers "as basic classifiers" and combines skills in a way that makes them better. Meta establishment adds to the LGBM classifier to improve final forecasts and increase classification projects.

B) “Dataset Collection”

"URL-based phishing datasets" are a lot of data that helps researchers and developers create a system that allows them to recognize the difference between phishing and real URLs. Kaggle, the Data Science Competition and Data Records website, comes from Kaggle.

“This is a general description of the dataset:”

“Name: Dataset for URL-based Phishing”

“Source: Kaggle”

“Purpose:” To make it easier to investigate and create systems that can spot phishing.

Size: “Has information from more than 11,000 websites.”

“Format: it is shown in vector form, which means that each URL is probably shown as a set of characteristics or attributes.”

It looks like a data file is set, so each item or instance is connected to the URL. Each URL has attributes (in vector form), which can use machine learning models to estimate whether the URL or the phishing page is an address.

The common elements in the Phishing detection data set could be things “like the length of the URL, the presence of specific keywords, the use of HTTPS, the age of domain and other related characters.” These features are very important for teaching ml models, how to determine the difference between real and phishing addresses 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

C) “Pre-processing”

“Using Pandas Data frame:” in this stage, we use Pandas, a powerful Python package for working with data, to clean, change, and get the dataset ready. this means dealing with missing numbers, changing the categories of data, and organizing the data so that it may be analyzed or modeled further.

“Using Seaborn and Matplotlib, we make charts and graphs to help us understand the features of the dataset.” This process helps us see patterns, correlations, and distributions in the data, which helps us make smart choices for the next study.

“*Label Processing:”* We use a preprocessing method called a label encoder to turn categorical labels into numbers. this is very important for ML models because they usually need numbers as inputs. “Label processing makes ensuring that the models can correctly read and learn from the dataset's categorical information.”

“*Feature selection:”* in this stage, we pick out the most important features from the data collection. feature selection is important for making models work better by focusing on the variables that give the most information and getting rid of noise. you can use statistical tests, correlation analysis, or ML methods to find the attributes that make the model's predictions more accurate.

D) “Training & Testing”

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

In the first part of our study, we used the first ML model (model 9) to look at and make sense of the preprocessed dataset. The extended phase then attempted to create predictions more accurately by creating hybrid models that combine predictions from multiple models. This new method attempts to make predictions more accurate, using the best parts of the various models. At the same time, we created an easy-to-use front-end based frontend, creating an authentication function that makes it easier for users to interact with the model. This makes it easier for users to enter data and receive predictions, making experience convenient and easy to use. The main goal of the project is to train the previously described ML models on previously processed data records so that they can find complex patterns and relationships within the data. After the training phase, the rigorous test is performed on a separate test data record. "We carefully use performance indicators like accuracy, precision, recall, and F1-score to see how well these algorithms can find phishing URLs." This thorough evaluation procedure is an important stage in quality assurance since it makes sure that the models are not only accurate but also reliable, which proves that they can be used in the actual world. Our initiative seeks to provide innovative and reliable solutions for detecting phishing URLs through this thorough technique.

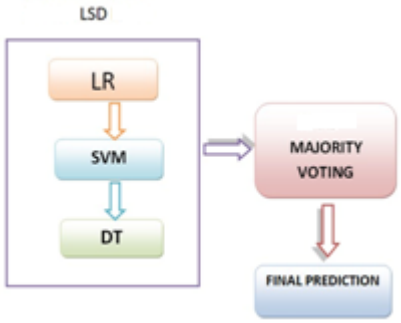
E) Algorithms.

"Stacking Classifier:"

The project uses the stacking classifier, which is the file method, to combine the prediction of "the random forest classifier [4] and the MLP classifier as the Foundation's classifiers. It uses the LGBM classifier as a meta-testimator to create a final prediction," which improves the ability of the project to classify things.

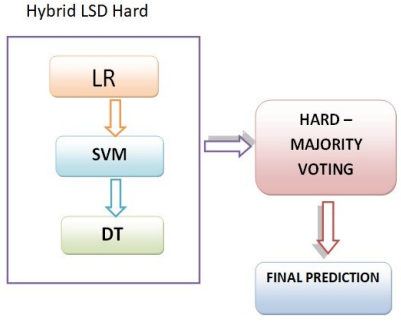
"LSD:"

"The LSD model with Hyperparameter (Logistic Regression, support Vector machine, decision Tree [4]) GridCV is a hybrid classification model that incorporates the best parts of the Logistic Regression, support Vector machine, and decision Tree [4]" methods to make it more accurate and faster. GridCV methodically goes through several combinations of hyperparameters to find the best ones for improving model performance. This makes it useful for a wide range of classification problems.



"Hybrid LSD (Hard):"

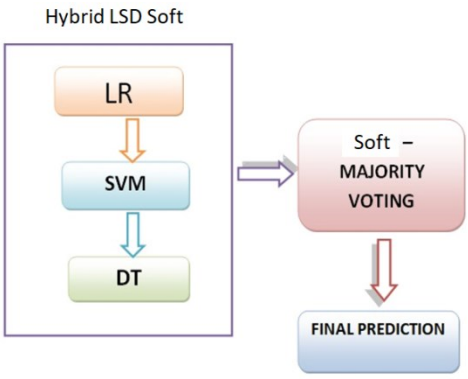
The "Hybrid LSD (hard) model uses the hard voting method along with the Logistic Regression, support Vector machine, and decision Tree [4] algorithms to make judgments about classification. each component model makes a prediction, and the final choice is reached by majority vote." This makes the model more accurate and reliable for a variety of classification tasks.



INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

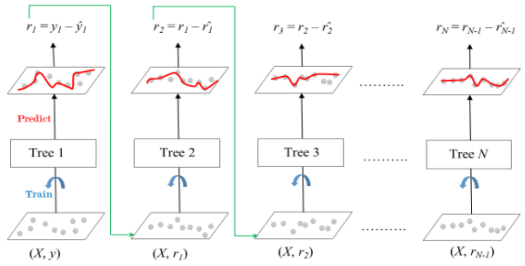
"Hybrid LSD (Soft):"

Using "soft voting to sort data, the Hybrid LSD (soft) model combines Logistic Regression, support Vector machine, and decision Tree [4]." It uses the best features of each model to produce predictions and can work with diverse types of data to make categorization tasks more accurate.



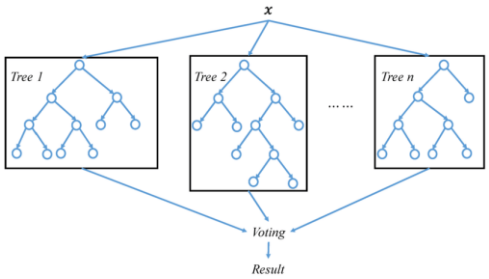
"Gradient Boosting:"

Graduate cavities are a type of ensemble-ML that creates predictive models step by step by step by step by step by step by combining some weak learner strengths, trees that usually make decisions. This is done by examining errors in previous models and modifying predictions. Ultimately, this is a very powerful and accurate prediction model, suitable for many tasks such as regression and classification.



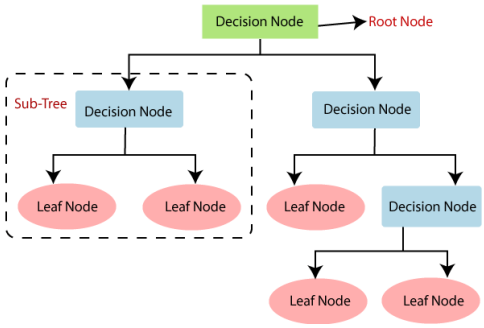
"Random Forest:"

Random Forest [4] is a way to learn from groups of decision groups [4] to create predictions. To make predictions, we train many decisions on average with random data groups [4]. This ensemble method improves accuracy, reduces the risk of excessive adaptation, and is suitable for both classification and regression problems.



"Decision Tree:"

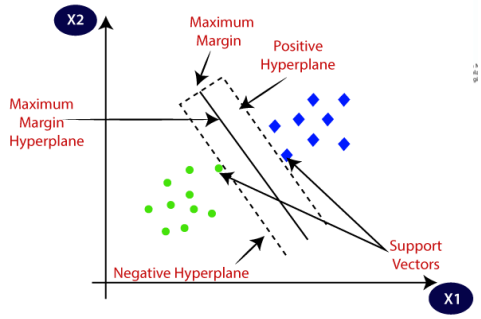
The decision tree [4] is a type of ML model that makes judgments repeatedly by distributing data into smaller groups depending on the most important feature. Its aim is to sort data or guess what happens next. It creates a structure similar to a tree where each node means function and each branch means a possible decision. This makes it easier to understand and useful for many tasks.



"Support Vector Classifier:"

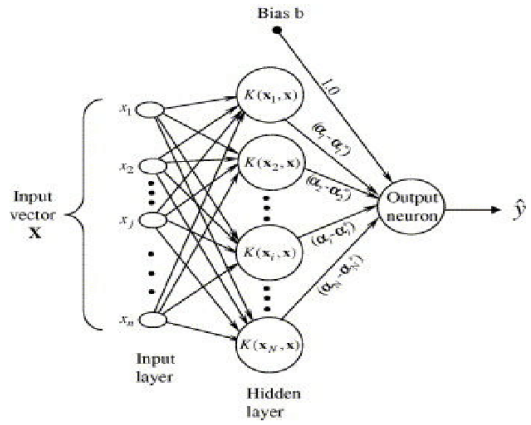
A "support Vector Classifier (SVC)" It is a type of ML model that identifies the best hyperveil, shares various data classes, and at the same time keeps the room as wide as possible. Find the most important support vectors for achieving correct classification. This is useful for both binary and classification applications with several classes.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025



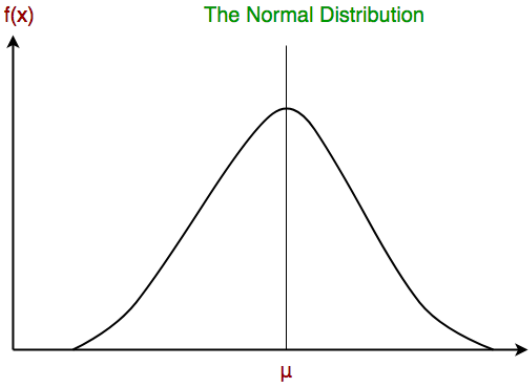
"Logistic Regression:"

Logistic regression is a kind of classification method that doubts how possible an input may belong to a particular category. Use the sigmoid function to convert the input function to a probability rating between 0 and 1. Then, depending on this score, it uses a threshold to assign the input into one of or more categories. during training, the model learns coefficients that help it fit the data and make correct classifications.



"Naive Bayes:"

"Naive Bayes is a type of probabilistic classification algorithm that uses Bayes' theorem and the "naive" idea that features are independent." based on the probabilities of its different properties, it figures out how likely it is that a data point belongs to a certain class. Naive Bayes works best for text categorization, spam detection, and other problems where it's k to assume that features are independent.



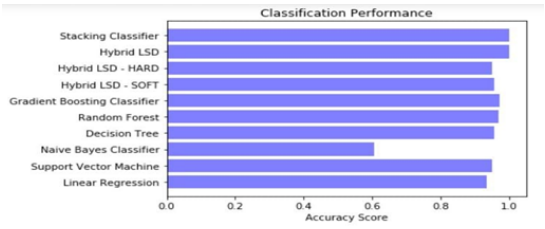
IV. EXPERIMENTAL RESULTS

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

"Accuracy:" Test accuracy is how accurate you can recognize the difference between weak and powerful examples. To measure how accurate the test is, a small number of actual positive and actual negative results should be monitored when thoroughly considered. This can be created with the following numbers:

"Accuracy = TP + TN TP + TN + FP + FN."

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



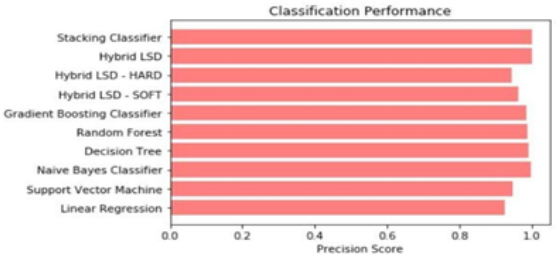
"Fig 2: Accuracy Graph"

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

"Precision:" Precision tells you how many of the positives were correctly categorized events or samples. So, you may use the following formula to figure out the accuracy:

"Precision = True positives/ (True positives + False positives) = TP/(TP + FP)"

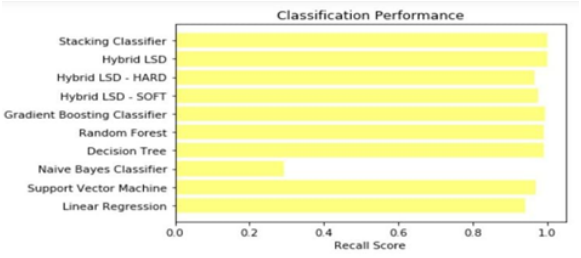
$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$



"Fig 3: Precision Score Graph"

"Recall:" A callback is an ML metric that measures how well a model can identify all related instances of a particular class. The ratio of correctly predicted positive impressions to actual positive outcomes indicates how well the model can catch a particular type of example.

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

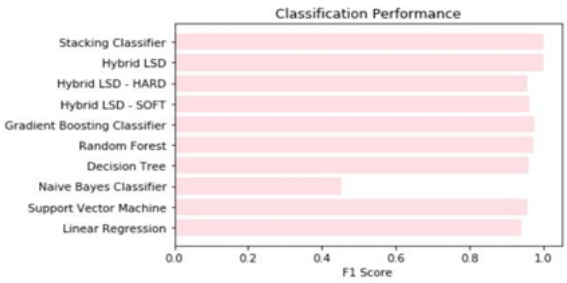


"Fig 4: Recall Score Graph"

"F1-Score: The F1 score is a way to measure how well a ML model works." It combines the review and precision scores of a model. The precision measurement tells you how often a model made the right prediction over the whole 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	0.605	0.454	0.292	0.997	0.909
3	Decision Tree	0.957	0.962	0.991	0.993	0.909
4	Random Forest	0.969	0.972	0.993	0.990	0.909
5	Gradient Boosting Classifier	0.974	0.977	0.994	0.986	0.909
6	Hybrid LSD - SOFT	0.959	0.964	0.977	0.965	0.909
7	Hybrid LSD - HARD	0.950	0.956	0.967	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"

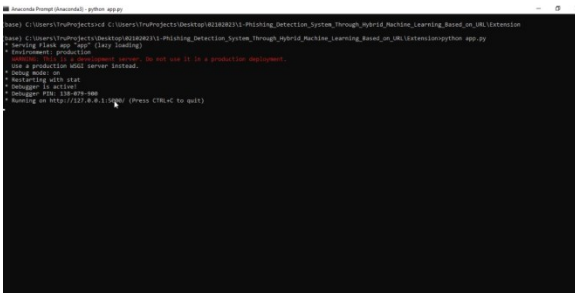
C) "Frontend"

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

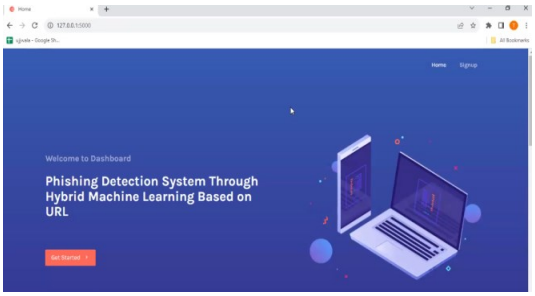
ISSN 2454-9940

www.ijasem.org

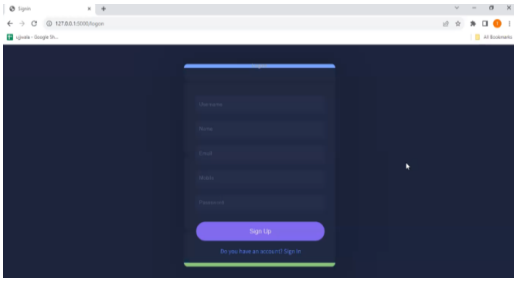
Vol 19, Issue 3, 2025



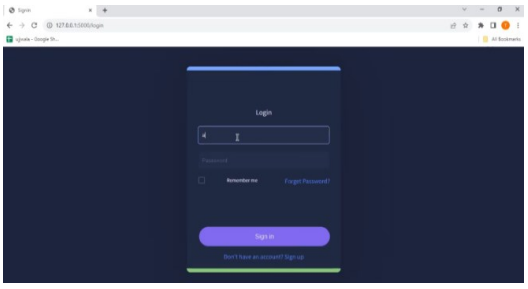
"Fig 7: Url Link to Web Page"



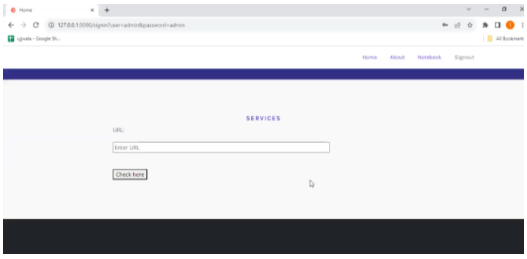
"Fig 8: Home page"



"Fig 9: User Signup page"

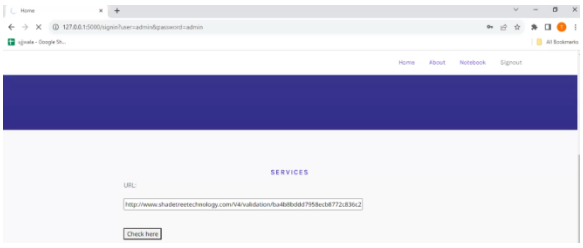


"Fig 10: User Sign in Page"

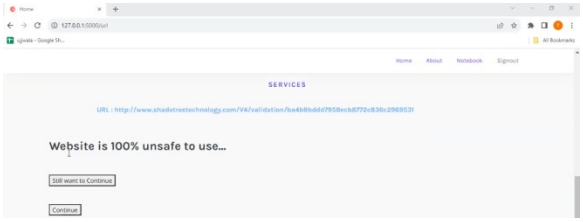


"Fig 11: Enter URL"

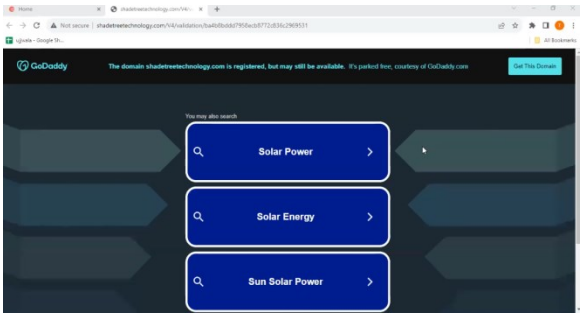
"Fig 12: Sample data for testing"



"Fig 13: Entered Url"

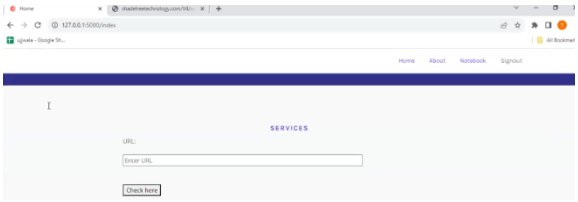


"Fig 14: Url result unsafe 100%"



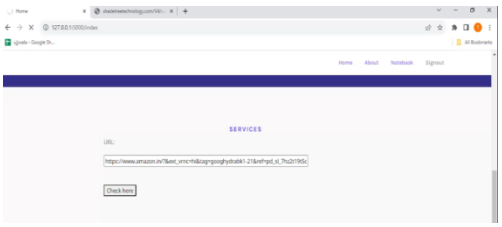
"Fig 15: Search Other Urls too"

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 3, 2025

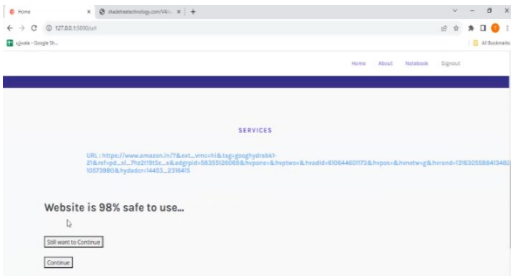


"Fig 15: Enter New Url"

"Fig 16: Sample data for testing"



"Fig 17: Entered New Url"



"Fig 18: Url result page (safe/ unsafe)"

V. CONCLUSION

In the end, the team used a hybrid ML technique that focused on URL properties to find phishing sites. The system become a lot more accurate and efficient by "using several models like decision Tree [4]s, Random forest [4]s, support vector classifiers, and an LSD-based stacking classifier." the choice of an extension stacking classifier stood out because of its "very high accuracy and F-score. This made the phishing detection system much more successful overall." This all-encompassing method solves a major problem in cybersecurity by offering strong protection against serious phishing assaults. adding more ML models not only made the system more flexible, but it also made it better able to respond to new phishing methods. The project's effectiveness in improving accuracy and efficiency shows how it could help improve cybersecurity measures, which would be a big help in the fight against cyber attacks. As phishing attempts get smarter, the system that was made is a strong defense mechanism that shows how it might be used in the real world to protect sensitive information and lower the risks that come with cyber threats.

VI. FUTURE SCOPE

The project's future goals include ongoing improvement and adjustment to new phishing methods. more study could look into how to combine DL, "behavioral analysis, and real-time threat intelligence to make the system's proactive defense even better. working with cybersecurity professionals" and people in the area can also help make the solution more complete and strong. This system will be more useful if it can be used in cloud environments and IoT devices, as well as have easy-to-use interfaces. The model will be effective since it will be updated regularly to reflect new threats. This makes it a solution in the ever-changing realm of cybersecurity.

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