ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 3, 2024

HYBRID MACHINE LEARNING-BASED PHISHING DETECTION SYSTEM USING URL

K BHASKAR[1], SOWMYA KARINGALPATTU[2], B AJITH KUMAR[3], K KAVITHA[4]
[1]Associate Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: bhaskark.mca@gmail.com
[2]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: ksowmya.20005@gmail.com
[3]Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: ajithkumaryadav34@gmail.com
[4]Assistant Professor, Department of MCA, Madanapalli Institute of Technology and Science, Email:kavithagoud32@gmail.com

Abstract: Phishing attacks are a serious internet based wrongdoing. Project utilizes phishing URL dataset. These URLs are involved by programmers for phishing. A few ML strategies are utilized to recognize and shut down these phishing endeavors. Decision tree, linear regression, random forest, naive Bayes, gradient boosting, K-neighbors, support vector, and LSD hybrid models are models. The task utilizes these calculations and refined strategies like cross-overlap validation and Grid Search Hyperparameter Optimization. The undertaking utilizes measurements to survey model execution. These estimations are precision, accuracy, recall, and F1-score. A LightGBM-based Stacking Classifier with RF + MLP further develops Phishing Detection System execution.

***Index terms** - Voting classifier, ensemble classifier, machine learning, uniform resource locator (URL), logistic regression, support vector machine, and decision tree (LSD), protocol, cyber security, social networks.*

1. INTRODUCTION

The web is crucial to many pieces of life. Web is an organization of PCs associated through telephone, fiber optic, remote, and satellite organizations. An overall PC organization. The web recovers information from hosts and servers. They imparted through IP-TCP. Government doesn't claim the Web; different associations, research offices, and colleges oversee it. Our happiness, instruction, finance, industry, web outsourcing, interpersonal interaction, clinical, and other everyday encounters have been made more straightforward by this. The web benefits numerous everyday issues.

Data search on the Web is great for instructive and research reasons. We may rapidly share information, motion pictures, pictures, and applications over email or create a letter to somebody all over the planet. The web utilizes online business. Web based business permits worldwide business and monetary exchanges. The 2020 Coronavirus pandemic has made internet based results more helpful for showing discoveries. Numerous web-based classes and conferences take time and are done on the web. Information sharing increments misfortune and cyberattack risk.

Web based buying is the fundamental Web application for vendors selling projects all around the world. Amazon has a monstrous internet based store. Web, used by Facebook, Instagram, WhatsApp, and other informal communities, makes fast and simple correspondence conceivable. Keeping a security strategy that safeguards correspondence and clients is essential. Aggressors can commit online misrepresentation, unsafe programming, PC infections, ransomware, worms, licensed innovation freedoms, refusal of administration attacks, illegal tax avoidance, defacing, electronic psychological warfare, and coercion on the Web.

Hacking, where anyone might take PC information and harm others, obliterates the Web. More youthful ages stress over corruption, which harms ethics. Individuals will profit from identifying these sites over basic, safe ones. Subsequently, these sites should be known. By contaminating a few PCs, infections can obliterate an organization and confidential information. It is improper to Utilize unlicensed sites. These parts require phishing identification to shield our PC framework. Network safety is an overall concern.

In the new ten years, different enemy of phishing detection techniques have been introduced. A unified resource locator (URL) structure in light of ML highlight determination approaches has been the subject of these works. The URL was made by Berners-Lee (1994). Previous sources and conventions lay out URL design. In 1985, space names with document way language structure were proposed. Slices isolated filenames and organizers from filepaths. Server names and document ways were isolated by twofold slices. To separate space names, Berners-Lee utilized specks. Progressively requested five parts make up HTTP URL linguistic structure.

2. LITERATURE SURVEY

A conventional province holds its actual qualities and its occupants' regular traditions, gifts, and social exercises. In any case, urbanization and monetary development are changing sure Malaysian customary networks. In this exploration, we distinguish the actual characteristics that are significant for social supportability in customary Malay settlements. This study utilized subjective techniques to portray Kuala Terengganu's customary settlements. Road configuration, house cutoff points, and open regions were displayed to protect social contact in three customary networks [1]. Consequently, the investigation discovered that actual characteristics and space typology choices are urgent to social practicality in conventional settlement gatherings.

Intrusion detection systems (IDS) should be sent across the organization to distinguish and hinder programmers' creating strategies. Each [3, 4] TCP/IP network layer has specific organization attacks, thus each needs a particular IDS. These days ML is the best device for network security since network-level information is monstrous and assault choices should be made rapidly and precisely. ML grouping can battle arising network intrusion threats. This part [4] utilizes Gaussian Naive Bayes, logistic regression, Decision Tree, and artificial neural networks on intrusion detection systems to distinguish typical and unusual TCP/IP attacks from publically open training datasets. We exhibit that Decision tree beat Gaussian in CoLab. Applying Naïve Bayes, Logistic regression, and Neural Network on a public dataset.

Because of its scale and variety, understanding the Internet's advancement is troublesome. We investigate roughly 1 trillion URLs questioned by a 2 million-man client board more than a year to concentrate on Web design and conduct [5]. URL lifetimes [5, 6, 13, 20] show that, not at all like early exploration, the assortment of URLs visited is very unique and all around displayed by a gamma circulation. We then inspect URL-crossing examples and find that hyperlink availability doesn't relate to perusing ways of behaving. This implies that hyperlink association doesn't straightforwardly influence client conduct Online. We finish up by talking about the way and question segments of URLs [32] and their properties in various site sorts. Because of these semantic qualifications, URL design might arrange the site it references [34, 36]. Our discoveries recommend URL standard alterations to advance Web the executives, straightforwardness, and semantic web progress.

Electronic exploration, particularly on HTML site pages because of their accessibility, has been endeavored for a really long time. The W3 consortium noticed that HTML [29] doesn't as expected portray the semantic design of website page contents in light of the fact that to its confined pre-characterized labels, semi-organized information, case awareness, and so on. Web engineers began utilizing XML and Glimmer to dodge these downsides. It empowers new examination approaches. XML URL arrangement in view of semantic primary direction is the subject of this paper [6]. The recommended strategy acquires 97.36% classification accuracy in tests.

This exploration inspects phishing and valid sites [7]. Our review improves phishing website detection. A ensemble learning approach in light of greater part casting a ballot is utilized with a feature selection calculation and contrasted with order models like Random forest, Logistic Regression, Prediction model, and so on. Our examination shows that current phishing detection techniques are 70%-92.52 percent exact. Exploratory discoveries demonstrate the way that our proposed model can recognize phishing sites with 95% accuracy, which is more noteworthy than current innovation. Learning models utilized in the examination show that our model has a promising precision rate.

Phishers currently use Twitter, Facebook, and Foursquare to disperse fakes. Famous miniature writing for a blog network Twitter permits 140-character tweets. North of 100 million individuals post 200 million tweets day to day. Due to this huge data scattering, phishers have begun utilizing Twitter to spread phishing. It is challenging to recognize phishing on Twitter contrasted with messages because of the speedy spread of phishing joins in the organization, short satisfied, and URL jumbling [15, 16, 17]. Our constant Twitter phishing identification strategy called PhishAri. Twitter-explicit and URL-explicit characteristics are utilized to recognize phishing tweets. We utilize Twitter-explicit elements including tweet content length, hashtags, and specifies. Twitter likewise utilizes the tweeter's age, measure of tweets, and adherent followee proportion. Twitter-explicit and URL-based capacities recognize phishing tweets well. [8] We distinguish phishing tweets with 92.52% precision utilizing AI order. End-clients might use our framework by means of a straightforward Chrome program module. The module groups tweets as phishing or protected progressively. We show the way that we can recognize phishing tweets at party time with high exactness, speedier than public boycotts and Twitter's own grouch identification technique. [8] In a lab research, we immediately assessed PhishAri's ease of use and viability and found that purchasers like and find it agreeable to use, in actuality. We accept this is the principal constant, thorough, and helpful Twitter phishing detection system.

3. METHODOLOGY

i) Proposed Work:

URL properties are utilized to recognize phishing endeavors in the proposed hybrid ML framework. It further develops accuracy with decision trees, random forests, and other ML methods. Cross-fold validation and hyperparameter tuning help its phishing [15] URL detection, giving solid cyberdefense. The Phishing Detection System project likewise utilizes a Stacking Classifier blending Random Forest (RF) and Multilayer Perceptron (MLP) with LightGBM to improve phishing detection. An easy to use Flask system with SQLite mix has been worked to further develop client testing and true convenience. It incorporates secure information exchange and signin capabilities.

ii) System Architecture:

- The methodology starts with an assortment of URLs [15, 16, 17] sorted as phishing or genuine. ML models are prepared and tried on this dataset.
- Guarantee information quality and consistency by eliminating or dealing with invalid or missing qualities prior to handling.
- The course of component designing believers URLs into mathematical element vectors. URL qualities like space length, extraordinary characters, and more are mathematically addressed. These element vectors feed ML calculations.
- The dataset contains training and testing information. Preparing information trains ML models, though testing information assesses them.
- Phishing prediction models utilize a few ML techniques. These models incorporate decision trees, random forests, SVMs, and others. We added a stacking classifier to the undertaking. Phishing URL designs are advanced by each model [26].
- Trained machine learning might arrange URLs as phishing or genuine utilizing feature vectors. Saving these prepared models for detection system incorporation is conceivable.
- The framework assesses trained models on testing information utilizing measurements. These measurements measure how effectively calculations perceive phishing URLs [17] with few false positives and negatives.

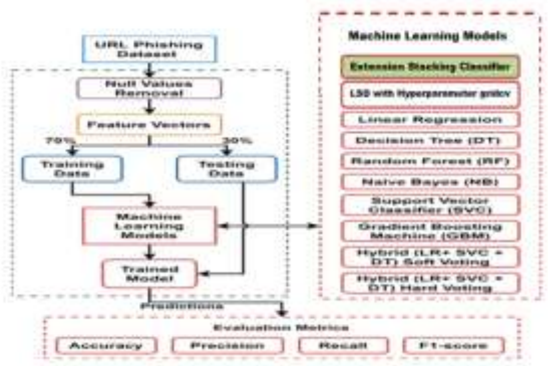
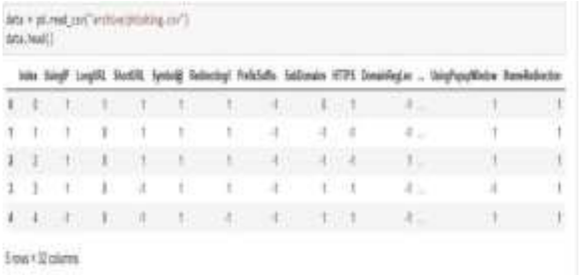


Fig 1 Proposed architecture

iii) Dataset collection:

Exploratory data analysis and feature relationship examination assist us with grasping Phishing URL Feature Data [27]. These strategies feature information appropriations, exceptions, and variable connections, empowering data processing and model development. The recommended strategy utilizes the "URL-based phishing dataset" from Kaggle, a famous dataset source. It contains vectorized phishing and lawful URLs from roughly 11,000 areas [15, 16, 17].



```
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
```

Fig 2 Dataset

The main 5 lines of the phishing data collection we assembled from Kaggle will be utilized to prepare the models. It has 32 segments, some of which are seen beneath.

iv) Data Processing:

Data processing transforms raw information into business-helpful data. Information researchers accumulate, sort out, clean, check, break down, and orchestrate information into diagrams or papers. Data can be handled physically, precisely, or electronically. Data ought to be more significant and decision-production simpler. Organizations might upgrade activities and settle on basic decisions quicker. PC programming improvement and other mechanized information handling innovations add to this. Big data can be transformed into significant bits of knowledge for quality administration and independent direction.

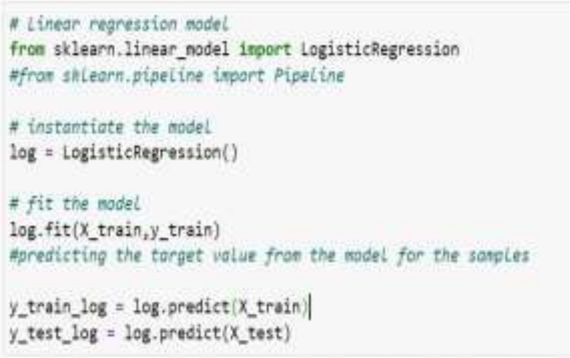
v) Feature selection:

Feature selection chooses the steadiest, non-repetitive, and pertinent elements for model turn of events. As data sets extend in amount and assortment, purposefully bringing down their size is significant. The fundamental reason for feature selection is to increment prescient model execution and limit processing cost.

One of the vital pieces of feature engineering is picking the main attributes for machine learning algorithms. To diminish input factors, feature selection methodologies take out copy or superfluous elements and limit the assortment to those generally critical to the ML model. Rather than permitting the ML model pick the main qualities, feature selection ahead of time enjoys a few benefits.

vi) Algorithms:

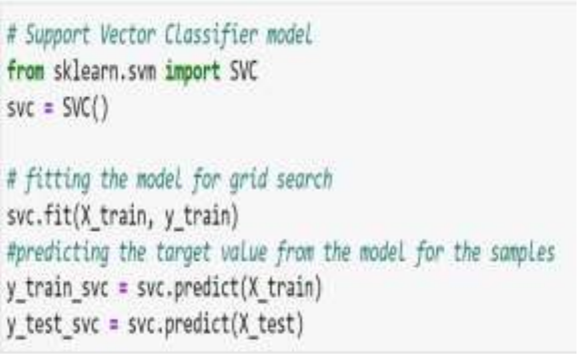
Logistic Regression is a classification technique that predicts input classification. It utilizes the sigmoid capability to move input qualities to a likelihood score somewhere in the range of 0 and 1, then, at that point, applies an edge to sort the contribution to at least two classifications. The model learns coefficients during preparing to fit information and group precisely.



```
# Linear regression model
from sklearn.linear_model import LogisticRegression
#from sklearn.pipeline import Pipeline

# instantiate the model
log = LogisticRegression()

# fit the model
log.fit(X_train,y_train)
#predicting the target value from the model for the samples

y_train_log = log.predict(X_train)
y_test_log = log.predict(X_test)
```

Fig 3 Linear regression

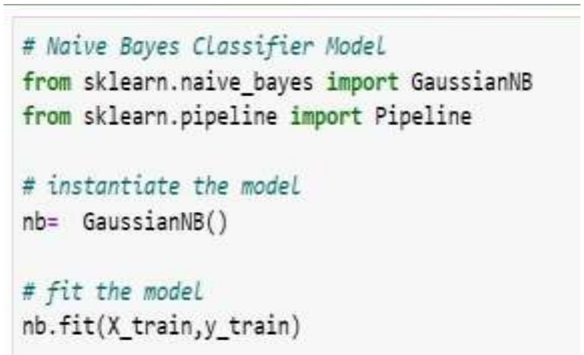
A Support Vector Classifier (SVC) is a ML model that recognizes the ideal hyperplane to divide information classes while expanding edge. The principal support vectors it sees as empower accurate binary and multi-class groupings.



```
# Support Vector Classifier model
from sklearn.svm import SVC
svc = SVC()

# fitting the model for grid search
svc.fit(X_train, y_train)
#predicting the target value from the model for the samples
y_train_svc = svc.predict(X_train)
y_test_svc = svc.predict(X_test)
```

Fig 4 SVC

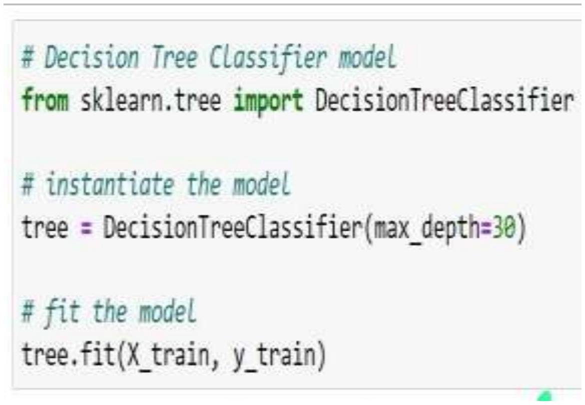
Naive Bayes is a probabilistic order procedure utilizing Bayes' hypothesis and "naive" feature freedom. It utilizes highlight probabilities to decide an information point's class. Naive Bayes succeeds in text arrangement, spam location, and other feature freedom based applications [9].



```
# Naive Bayes Classifier Model
from sklearn.naive_bayes import GaussianNB
from sklearn.pipeline import Pipeline

# instantiate the model
nb=  GaussianNB()

# fit the model
nb.fit(X_train,y_train)
```

Fig 5 Naïve bayes

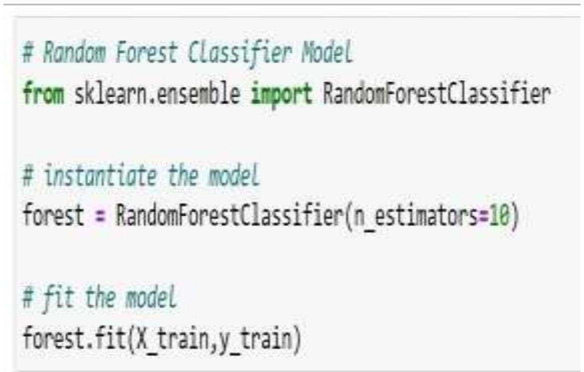
A Decision Tree is a ML model that characterizes or predicts results by recursively isolating information into subgroups relying upon the main trait. It constructs a tree-like design with hubs addressing features and branches addressing elective choices, making it interpretable and gainful for different errands.



```
# Decision Tree Classifier model
from sklearn.tree import DecisionTreeClassifier

# instantiate the model
tree = DecisionTreeClassifier(max_depth=30)

# fit the model
tree.fit(X_train, y_train)
```

Fig 6 Decision tree

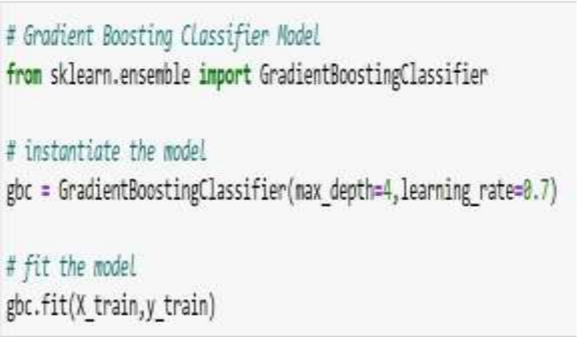
Random Forest is an ensemble learning approach that predicts utilizing a few decision trees. Training decision trees on arbitrary information subsets and averaging their expectations works. This ensemble strategy further develops classification and regression accuracy, wipes out overfitting, and performs well.



```
# Random Forest Classifier Model
from sklearn.ensemble import RandomForestClassifier

# instantiate the model
forest = RandomForestClassifier(n_estimators=10)

# fit the model
forest.fit(X_train,y_train)
```

Fig 7 Random forest

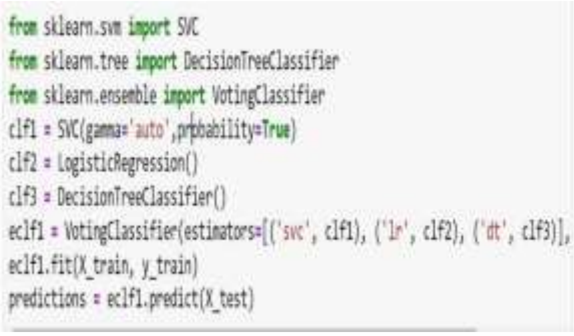
Gradient Boosting is an ensemble ML technique that steadily fosters a prescient model utilizing frail students, for the most part decision trees. It achieves so by zeroing in on earlier model blames and changing its expectations to diminish them, producing a strong and exact prescient model that succeeds in regression and classification.



```
# Gradient Boosting Classifier Model
from sklearn.ensemble import GradientBoostingClassifier

# instantiate the model
gbc = GradientBoostingClassifier(max_depth=4,learning_rate=0.7)

# fit the model
gbc.fit(X_train,y_train)
```

Fig 8 Gradient boosting

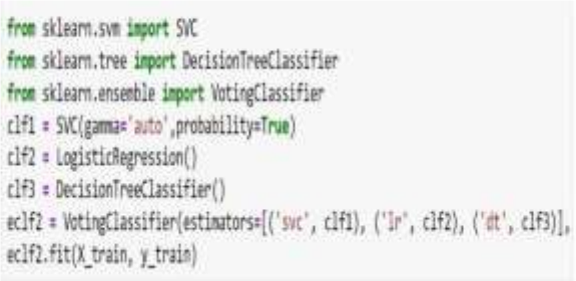
The Hybrid LSD (Soft) model uses soft voting, logistic regression, support vector machine, and decision tree to classify data. It leverages the capabilities of each model to predict and process different data to improve classification accuracy.



```
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import VotingClassifier
clf1 = SVC(gamma='auto',probability=True)
clf2 = LogisticRegression()
clf3 = DecisionTreeClassifier()
eclf1 = VotingClassifier(estimators=[('svc', clf1), ('lr', clf2), ('dt', clf3)],
eclf1.fit(X_train, y_train)
predictions = eclf1.predict(X_test)
```

Fig 9 Hybrid LSD (soft

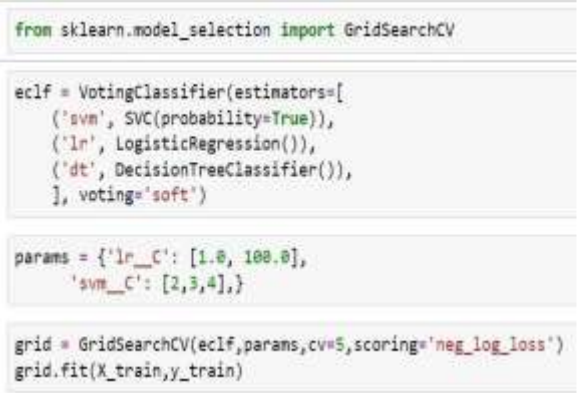
The Hybrid LSD (Hard) model uses logistic regression, support vector machines, decision tree techniques, and hard voting for classification. Each component model makes a prediction and the final decision is made by majority vote, which improves classification accuracy and stability.



```
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import VotingClassifier
clf1 = SVC(gamma='auto',probability=True)
clf2 = LogisticRegression()
clf3 = DecisionTreeClassifier()
eclf2 = VotingClassifier(estimators=[('svc', clf1), ('lr', clf2), ('dt', clf3)],
eclf2.fit(X_train, y_train)
```

Fig 10 Hybrid LSD (Hard)

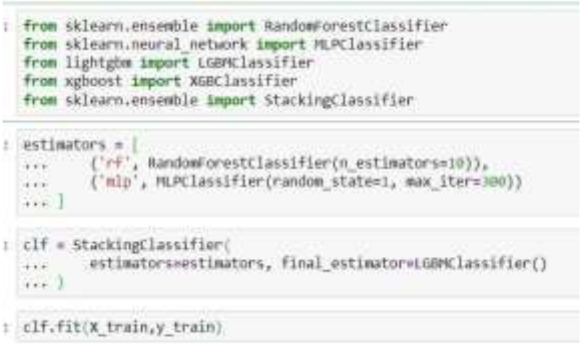
The LSD (Logistic Regression, Support Vector Machine, Decision Tree) GridCV hyperparametric models are hybrid classification models that combine logistic regression, support vector machine, and decision tree techniques to improve accuracy and efficiency. GridCV optimizes model performance by searching combinations of hyperparameters, making it useful for classification problems.



```
from sklearn.model_selection import GridSearchCV

eclf = VotingClassifier(estimators=[
    ('svm', SVC(probability=True)),
    ('lr', LogisticRegression()),
    ('dt', DecisionTreeClassifier()),
    ], voting='soft')

params = {'lr__C': [1.0, 100.0],
      'svm__C': [2,3,4],}

grid = GridSearchCV(eclf,params,cv=5,scoring='neg_log_loss')
grid.fit(X_train,y_train)
```

Fig 11 LSD with Hyperparameter GridCV

The project uses Ensemble StackingClassifier to combine the predictions of RandomForestClassifier and MLPClassifier as base classifiers. LGBMClassifier is used as a meta-estimator to predict the final result, improving classification performance.



```
from sklearn.ensemble import RandomForestClassifier
from sklearn.neural_network import MLPClassifier
from lightgbm import LGBMClassifier
from xgboost import XGBClassifier
from sklearn.ensemble import StackingClassifier
```

```
estimators = [
...     ('rf', RandomForestClassifier(n_estimators=10)),
...     ('mlp', MLPClassifier(random_state=1, max_iter=300))
... ]
```

```
clf = StackingClassifier(
...     estimators=estimators, final_estimator=LGBMClassifier()
... )
```

```
clf.fit(X_train,y_train)
```

Fig 12 Stacking classifier

4. EXPERIMENTAL RESULTS

Precision: Precision estimates the level of positive cases or tests precisely sorted. Precision is determined utilizing the recipe:

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

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

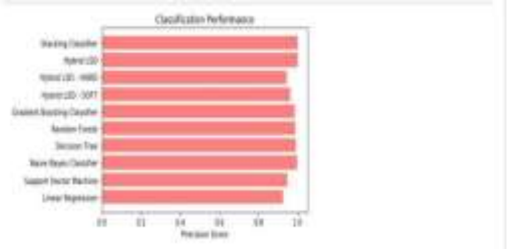


Fig 13 Precision comparison graph

Recall: Machine learning recall assesses a model's ability to perceive all significant examples of a class. It shows a model's culmination in catching occasions of a class by contrasting accurately anticipated positive perceptions with complete positives.

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

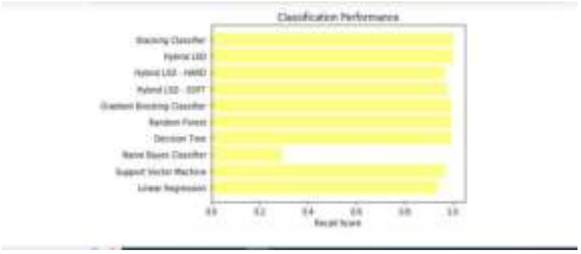


Fig 14 Recall comparison graph

Accuracy: A test's accuracy is its ability to recognize debilitated from sound cases. To quantify test accuracy, figure the small part of true positive and true negative in completely broke down cases. Numerically, this is:

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

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

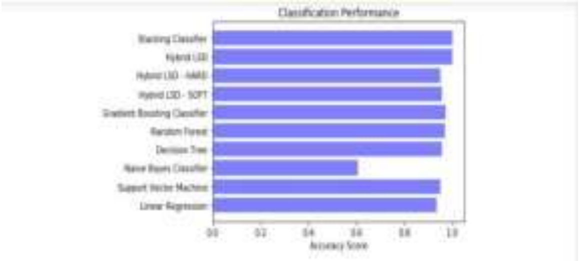


Fig 15 Accuracy graph

F1 Score: Machine learning model accuracy is estimated by F1 score. Consolidating model precision and recall scores. The accuracy measurement estimates how frequently a model anticipated accurately all through the dataset.

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

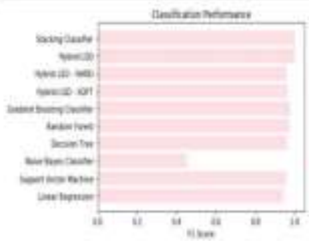
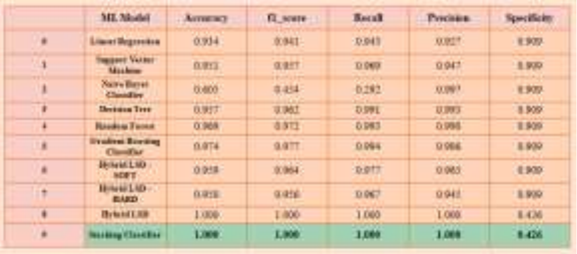


Fig 16 F1Score

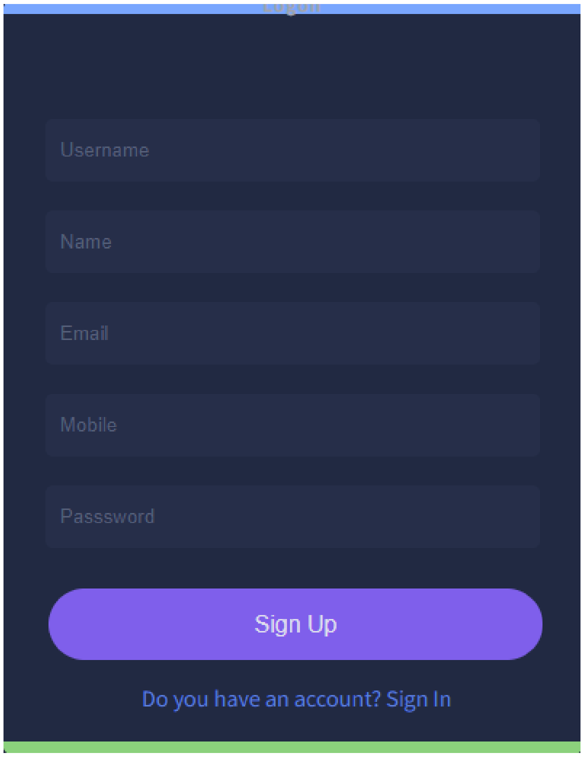


	ML Model	Accuracy	f1_score	Recall	Precision	Specificity
0	Linear Regression	0.934	0.941	0.943	0.927	0.909
1	Support Vector Machine	0.953	0.957	0.969	0.947	0.909
2	Naive Bayes Classifier	0.600	0.454	0.282	0.997	0.909
3	Decision Tree	0.957	0.962	0.991	0.993	0.909
4	Random Forest	0.968	0.972	0.983	0.988	0.909
5	Gradient Boosting Classifier	0.974	0.977	0.994	0.986	0.909
6	Hybrid LSD - SOFT	0.958	0.964	0.977	0.983	0.909
7	Hybrid LSD - HARD	0.958	0.956	0.967	0.945	0.909
8	Hybrid LSD	1.000	1.000	1.000	1.000	0.436
9	Stacking Classifier	1.000	1.000	1.000	1.000	0.426

Fig 17 Performance Evaluation

Phishing Detection System Through Hybrid Machine Learning Based on URL

Fig 18 Home page



Login

Username

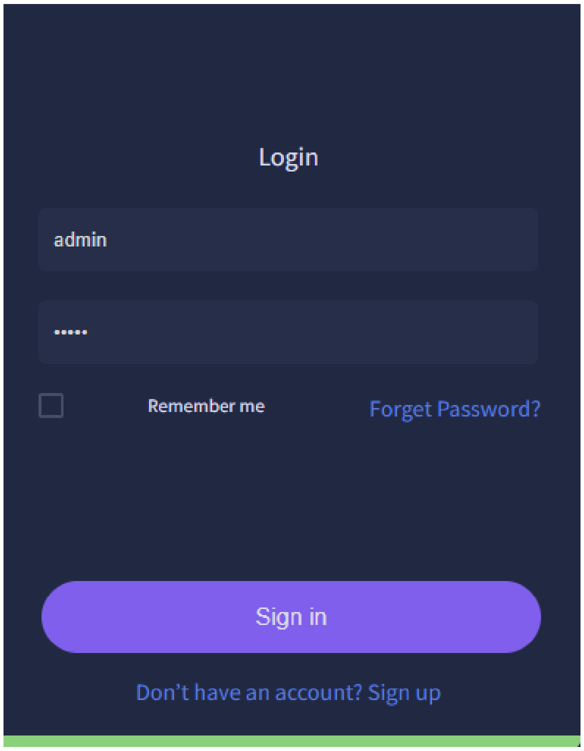
Name

Email

Mobile

Passsword

Fig 19 Signin page



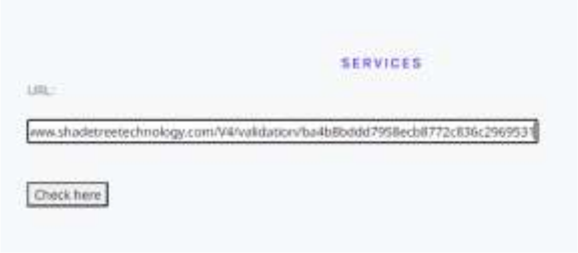
Login

admin

•••••

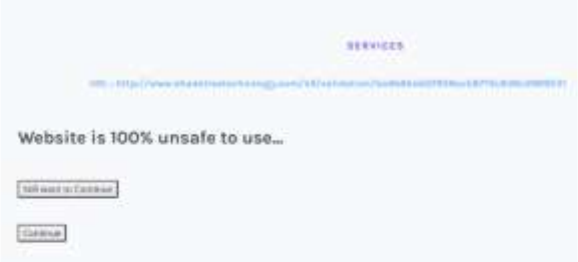
□ Remember me

Fig 20 Login page



www.shadetreetechnology.com/V4/validation/ba4b8bddd7958ecb8772c836c2969531

Fig 21 User input



Website is 100% unsafe to use...

Fig 22 Predict result for given input

5. CONCLUSION

The review utilized hybrid machine learning to improve phishing detection by underscoring URL properties. Decision tree, random forest, support vector classifier, LSD (both hard and soft), stacking classifier technique augmentation, and Mixture LSD, all with amazing accuracy and F1-score, were utilized to further develop the framework's phishing threat detection [2, 3]. Moreover, a stacking classifier was picked for its high accuracy and F-score, working on the framework's exhibition. This decision shows devotion to a high-performing phishing detection calculation. Flask and SQLite give a protected and smooth front-end for client testing, working on the framework's ease of use. Client information exchange, signin, and testing are smooth with this combo. The undertaking's complete phishing detection system safeguards against serious phishing attacks, a basic network protection issue [7, 8]. The task's utilization of cutting edge ML, model assortment, and easy to understand highlights shows its obligation to tackling cybersecurity challenges in the evolving scene.

6. FUTURE SCOPE

Phishing detection might be improved by investigating new calculations and techniques [11, 12, 13, 16]. The framework might expect new risks by examining and executing new techniques. Due to changing digital risks, ongoing checking should be incorporated to the venture. The framework can quickly adjust to new go after strategies and furnish opportune insurance with this proactive methodology. To check model solidness and generalizability, survey it on a greater and more shifted dataset. This drawn out dataset better addresses certifiable conditions, further developing framework execution. Incorporating client conduct and network traffic analysis helps boost system detection. The framework can investigate dangers all the more precisely by utilizing more information. Past phishing, the methodology can distinguish malware and social designing. This widened broadness allows the framework actually to battle greater online protection dangers.

REFERENCES

[1] N. Z. Harun, N. Jaffar, and P. S. J. Kassim, ‘‘Physical attributes significant in preserving the social sustainability of the traditional malay settlement,’’ in Reframing the Vernacular: Politics, Semiotics, and Representation. Springer, 2020, pp. 225–238.

[2] D. M. Divakaran and A. Oest, ‘‘Phishing detection leveraging machine learning and deep learning: A review,’’ 2022, arXiv:2205.07411.

[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] J. Kline, E. Oakes, and P. Barford, ‘‘A URL-based analysis of WWW structure and dynamics,’’ in Proc. Netw. Traffic Meas. Anal. Conf. (TMA), Jun. 2019, p. 800.

[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.

[31] C. L. Tan, K. L. Chiew, K. Wong, and S. N. Sze, ‘‘PhishWHO: Phishing webpage detection via identity keywords extraction and target domain name finder,’’ Decis. Support Syst., vol. 88, pp. 18–27, Aug. 2016.

[32] A. Le, A. Markopoulou, and M. Faloutsos, ‘‘PhishDef: URL names say it all,’’ in Proc. IEEE INFOCOM, Apr. 2011, pp. 191–195.

[33] R. Islam and J. Abawajy, ‘‘A multi-tier phishing detection and filtering approach,’’ J. Netw. Comput. Appl., vol. 36, no. 1, pp. 324–335, Jan. 2013.

[34] S. C. Jeeva and E. B. Rajsingh, ‘‘Intelligent phishing URL detection using association rule mining,’’ Hum.-Centric Comput. Inf. Sci., vol. 6, no. 10, pp. 1–19, 2016.

[35] M. Babagoli, M. P. Aghababa, and V. Solouk, ‘‘Heuristic nonlinear regression strategy for detecting phishing websites,’’ Soft Comput., vol. 23, no. 12, pp. 4315–4327, Jun. 2019.

[36] E. Buber, B. Diri, and O. K. Sahingoz, ‘‘Detecting phishing attacks from URL by using NLP techniques,’’ in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 337–342.

[37] E. Buber, B. Diri, and O. K. Sahingoz, ‘‘NLP based phishing attack detection from URLs,’’ in Proc. Int. Conf. Intell. Syst. Design Appl. Cham, Switzerland: Springer, 2017, pp. 608–618.

[38] R. M. Mohammad, F. Thabtah, and L. McCluskey, ‘‘Predicting phishing websites based on self-structuring neural network,’’ Neural Comput. Appl., vol. 25, no. 2, pp. 443–458, Aug. 2014.

[39] F. Feng, Q. Zhou, Z. Shen, X. Yang, L. Han, and J. Wang, ‘‘The application of a novel neural network in the detection of phishing websites,’’ J. Ambient Intell. Hum. Comput., vol. 14, pp. 1–15, Apr. 2018.

[40] S. Smadi, N. Aslam, and L. Zhang, ‘‘Detection of online phishing email using dynamic evolving neural network based on reinforcement learning,’’ Decis. Support Syst., vol. 107, pp. 88–102, Mar. 2018.

[41]G.Viswanath, “Hybrid encryption framework for securing big data storage in multi-cloud environment”, Evolutionary intelligence, vol.14, 2021, pp.691-698.

[42] Viswanath Gudditi, “Adaptive Light Weight Encryption Algorithm for Securing Multi-Cloud Storage”, Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol.12, 2021, pp.545-552.

[43] Viswanath Gudditi, “A Smart Recommendation System for Medicine using Intelligent NLP Techniques”, 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), 2022, pp.1081-1084.

[44] G.Viswanath, “Enhancing power unbiased cooperative media access control protocol in manets”, International Journal of Engineering Inventions, 2014, vol.4, pp.8-12.

[45] Viswanath G, “A Hybrid Particle Swarm Optimization and C4.5 for Network Intrusion Detection and Prevention System”, 2024, International Journal of Computing, DOI: https://doi.org/10.47839/ijc.23.1.3442, vol.23, 2024, pp.109-115.

[46] G.Viswanath, “A Real Time online Food Ording application based DJANGO Restfull Framework”, Juni Khyat, vol.13, 2023, pp.154-162.

[47] Gudditi Viswanath, “Distributed Utility-Based Energy Efficient Cooperative Medium Access Control in MANETS”, 2014, International Journal of Engineering Inventions, vol.4, pp.08-12.

[48] G.Viswanath,“ A Real-Time Video Based Vehicle Classification, Detection And Counting System”, 2023, Industrial Engineering Journal, vol.52, pp.474-480.

[49] G.Viswanath, “A Real- Time Case Scenario Based On Url Phishing Detection Through Login Urls ”, 2023, Material Science Technology, vol.22, pp.103-108.

[50] Manmohan Singh,Susheel Kumar Tiwari, G. Swapna, Kirti Verma, Vikas Prasad, Vinod Patidar, Dharmendra Sharma and Hemant Mewada, “A Drug-Target Interaction Prediction Based on Supervised Probabilistic Classification” published in Journal of Computer Science, Available at: https://pdfs.semanticscholar.org/69ac/f07f2e756b79181e4f1e75f9e0f275a56b8e.pdf