IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

IS IT PHISHING OR NOT? A SURVEY ON PHISHING WEBWEBSITE DETECTION

T ANIL KUMAR[1], P VINODH[2], K BHASKAR[3], N SRIDEVI[4]
[1]Associate Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: anil.thumburu@gmail.com
[2]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: pothurajuvinod912@gmail.com
[3]Associate Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: bhaskark.mca@gmail.com
[4]Assistant Professor, Department of CSE, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: sreepsrl@gmail.com

Abstract: Phishing is a determined and fruitful security issue that compromises people and designated brands, underlining the need for compelling detection and insurance. Complete Detection Techniques Review**: Phishing webwebsite detection will be totally evaluated in the review. It looks to make sense of recognition strategies and their viability. A thorough assessment of rundown based, comparability based, and ML-based detection methods is finished. It additionally looks at the datasets used to evaluate different procedures, uncovering their assets and shortcomings. The venture features phishing webwebsite detection research holes that need more review and improvement to further develop detection strategies. The undertaking's Voting Classifier (MLP+XGB+Decision tree classifier) distinguishes phishing webwebwebsites better. An easy to use Flask framework with SQLite joining improves on user testing enlistment and signin, making network protection applications usable.

Index terms *- Phishing, security threat, phishing webwebsite, phishing detection, URL, blacklists, machine learning, page similarity, datasets, social engineering.*

1. INTRODUCTION

Phishing is a perilous security issue that utilizations complex mental and social designing to beguile individuals into clicking connections to pernicious webwebsites and submitting delicate data like individual or business data and record passwords. Phishing assaults [7, 8, 10, 14, 24] are easy to send. They're normally compelling. Aggressors foster very much planned phishing webwebwebsites that show up and feel like valid destinations, making them hard to recognize. As per [1], aggressors have fostered their techniques and avoidance procedures over the course of time to sidestep identification.[51]

Phishing assaults make immediate and roundabout impacts. Phishing might think twice about and accounts, bringing about cash robbery and an emergency of confidence in web-based administrations. These attacks likewise hurt the organizations and associations being mimicked, who might endure information breaks, monetary misfortunes, and notoriety harm.

Enisa [2] found that European SME cyberattacks are most frequently phishing attacks. Cisco's Network safety danger patterns study [3] gauges that 90% of information breaks in 2020 will include phishing. No less than one individual attempted to interface with a phishing website in 86% of organizations. As depicted in [4], phishing attacks are normal since individuals don't as expected assess webwebsites and don't get sufficient training. The APWG Phishing movement patterns study [5] tracked down north of 1,000,000 phishing webwebsites in the main quarter of 2022.

Many investigations have inspected phishing website discovery over time. Our study surveys the most significant phishing webwebsite detection systems in the writing to give an exhaustive and complete evaluation.

Aggressors use typosquatting and combosquatting to make connections and webwebsites look legitimate to earn certainty. They make the Uniform Resource Locator (URL) designs in the program's detection bar by putting unessential accentuation marks (e.g., run), incorrectly spelled words (e.g., paymet), or specific terms (e.g., brand name being designated) in improper areas. Assailants might utilize comparable letters from different letter sets to substitute English ones. Albeit false pages might be facilitated on hacked servers, aggressors might enroll areas with real names. Because of the minimal expense of laying out new phishing URLs, assailants only sometimes reuse them, making phishing webwebsite detection a lot harder. Recollect that a URL [11, 12] is a comprehensible series of characters deciphered by client projects to exceptionally distinguish a web website [6].

2. LITERATURE SURVEY

Web clients are in danger from phishing. Phishing webwebsites are turning out to be more complex, and avoidance strategies could allow them to sidestep the environment's detection measures and cause true harm. Shrouding, a refined client-side avoidance technique that utilizes JavaScript to permit complicated collaborations between possible casualties and the phishing website, can help hinder or impede robotized alleviations. The rate and effect of client-side shrouding have not been examined.In [1], we offer CrawlPhish, a framework for naturally distinguishing and grouping client-side shrouding on known phishing webwebsites. We use CrawlPhish to gather and assess 112,005 wild phishing webwebsites more than 14 months in 2018-2019. Utilizing state of the art static and dynamic code examination, we uncover 1,128 client-side shrouding techniques on 35,067 webwebsites. Toward the finish of our information assortment period, aggressors' shrouding utilize expanded from 23.32% to 33.70%. Our system has 1.45% misleading positive and 1.75% bogus negative shrouding discovery rates. We characterize eight avoidance techniques into three undeniable level classes: Client Association, Fingerprinting, and Bot Conduct in light of the semantics of the methodologies we found. We show that every avoidance approach might sidestep program based phishing recognition (a key biological system security) utilizing 150 phony phishing webwebsites [30, 36, 38]. A client research affirms that the strategies don't deflect casualty visits. Subsequently, we recommend utilizing our procedure to build the environment's ability to forestall phishing webwebsites utilizing client-side shrouding and constantly distinguish assailant sent off shrouding systems.[53]

The twenty years prior and presently, phishing was a digital threat. Phishers' imaginative assault arranging and execution have created it over the long run. Hence, phishing history and strategies should be evaluated. Here is an efficient, far reaching, and straightforward survey of these techniques. Every technique's media and vectors are distinguished. Medium is the stage where the methodologies reside and vector is the means by which the phisher spreads the attack. Definite depiction of these strategies rules [7]. Likewise featured is the way phishers utilized these strategies. This survey will work on comprehension of existing phishing methodologies and assist with planning a complete enemy of phishing arrangement. This survey raises' comprehension perusers might interpret phishing systems and supports phishing aversion. Also, this assessment will direct research by phishing type and recognize regions where against phishing endeavors are absent. Policymakers and against phishing engineers will profit from this survey.

Web-based entertainment and internet banking have made individuals' life more straightforward thanks to the web. Framework and organization security gambles are continually developing because of Web innovations. Phishing [44, 45, 46, 47] is an extreme risk where assailants use sham messages or webwebsites to get client qualifications. Industry and scholastics are creating phishing arrangements. End-client mindfulness is vital to phishing danger evasion for associations. Our article has two objectives [8]. We'll begin with phishing's set of experiences and aggressors' inspirations. We will then, at that point, arrange phishing attacks. Second, utilizing our scientific categorization, we will group artistic answers for safeguard clients from phishing. We finish our exploration by referencing abstract worries and challenges that are important to counter phishing assaults.

Online monetary exchanges are high in the time of electronic and portable business, setting out misrepresentation open doors. Phishing is a regular trick that incorporates building a phony website to take client certifications. Website phishing costs banks, web clients, states, and different associations large chunk of change. Phishing might be battled by showing novice clients on phisher procedures through courses or preparing. Since phishing procedures change, this technique might be costly and wasteful. Regulation or change of network protection regulation that uphold online fraudsters' discipline is another enemy of phishing procedure. Shrewd ML innovation is a superior enemy of phishing system. Utilizing this method, the program coordinates a classification framework to distinguish phishing and inform clients. Lawful, preparing, instructive, and shrewd enemy of phishing techniques are seriously analyzed in this exploration [9]. Significantly, wise and customary phishing protections are differentiated, alongside their benefits and cons from a client and execution viewpoint. PC security subject matter experts, web security analysts, and company proprietors might benefit from this website phishing assessment.

Malignant URLs, or webwebsites, are a significant network safety issue. Malignant URLs have undesirable material (spam, phishing, drive-by takes advantage of, and so forth) and snare unwary clients to become trick casualties (financial misfortune, confidential data robbery, and malware establishment), costing billions of dollars yearly. Opportune identification and reaction to such dangers are fundamental. Boycotts are utilized for this discovery. Boycotts can't distinguish recently created malevolent URLs. Late years have seen expanded interest in ML ways to deal with improve vindictive URL recognition comprehensiveness. This article [11] will inspect and structure AI based malevolent URL detection strategies. We formalize Noxious URL Detection as an AI challenge and order and survey writing deals with on problem perspectives including highlight portrayal and calculation plan. Further, this article gives an opportune and thorough review for different crowds, remembering ML specialists and designers for the scholarly world and network safety experts and professionals [9, 45], to assist them with grasping the cutting edge and plan their own exploration and applications. We investigate framework plan reasonable items, open exploration challenges, and urgent future examination regions.

3. METHODOLOGY

i) Proposed Work:

Phishing detection is shrouded exhaustively all through the task, including list-based, similitude based, and ML calculations, datasets, issues, literary highlights, and human angles. It suggests utilizing ML and AI [11] to further develop detection, teaming up to share information, and underscoring the significance of schooling and mindfulness in phishing avoidance. With a perplexing Voting Classifier integrating MLP, XGBoost, and Decision Tree Classifier, the task improves phishing website detection. Utilizing fluctuated classifier qualities further develops execution in this group technique. The venture utilizes an easy to use Flask framework with SQLite to smooth out online protection application user testing enrollment and signin processes [9, 45]. This mix of strong detection strategies and a smooth UI improves phishing website detection.[55]

ii) System Architecture:

A multi-layered strategy including blacklisting, whitelisting, textual analysis, visual similarity comparison, and machine learning is used to detect questionable online sites [11]. Web sites should be correctly classified as legitimate or phishing. The architecture starts with online pages to be checked for phishing.

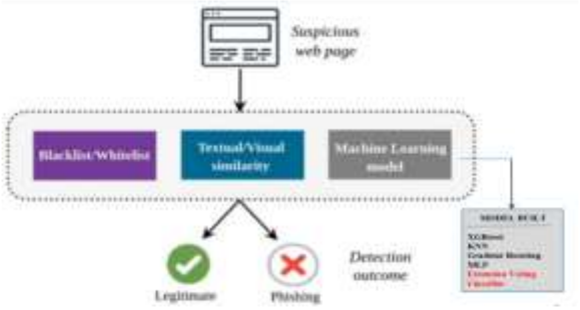
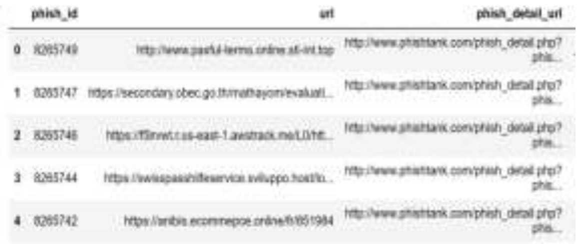


Fig 1 Proposed architecture

iii) Dataset collection:

Phishtank and Curlie datasets into your project. These datasets may contain URLs [11, 12], domains, and other phishing detecting elements. As said, blacklists and whitelists are generated using different methods that consider attacker and individual behavior. Phishing and authentic websites from diverse sources are used to evaluate these methods. PhishTank [35], a community-based phishing website reporting and verification system, and Google Safe Browsing lists feature common dangerous URL sources. Alexa, which departed in May 2022, and DMOZ, which closed in 2017 and was replaced by Curlie [36], were also suppliers of benign URLs.



	phish_id	url	phish_detail_url
0	8265749	http://www.pasful-terms.online.at-int.top	http://www.phishtank.com/phish_detail.php?phis...
1	8265747	https://secondary.obec.go.th/mathayom/evaluati...	http://www.phishtank.com/phish_detail.php?phis...
2	8265746	https://f5imwt.r.us-east-1.awstrack.me/L0/htt...	http://www.phishtank.com/phish_detail.php?phis...
3	8265744	https://swisspasshifleservice.sviluppo.host/lo...	http://www.phishtank.com/phish_detail.php?phis...
4	8265742	https://anbis.ecommepoe.online/h/851984	http://www.phishtank.com/phish_detail.php?phis...

Fig 2 NSL KDD dataset

iv) Data Processing:

Data processing turns raw data into business-useful information. Data scientists gather, organize, clean, verify, analyze, and arrange data into graphs or papers. Data can be processed manually, mechanically, or electronically. Information should be more valuable and decision-making easier. Businesses may enhance operations and make critical choices faster. Computer software development and other automated data processing technologies contribute to this. Big data can be turned into relevant insights for quality management and decision-making.

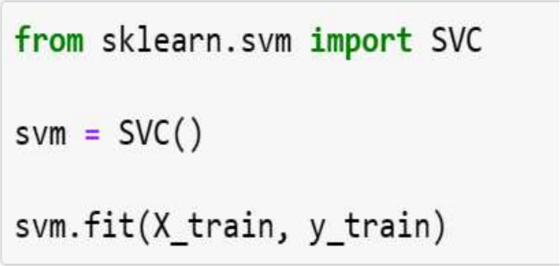
v) Feature selection:

Feature selection chooses the most steady, 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.[57]

vi) Algorithms:

Support Vector Machine (SVM) - A strong supervised learning technique for classification is SVM. It identifies the best hyperplane in high-dimensional space to classify data. SVM uses URL and webpage content information to categorize websites as phishing or real.



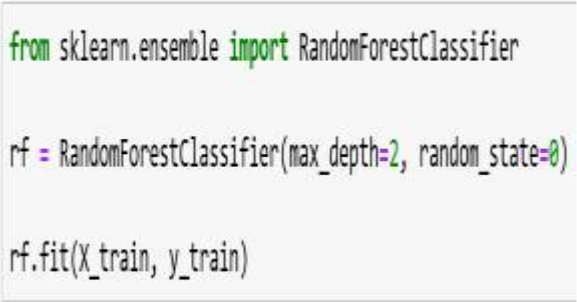
```
from sklearn.svm import SVC

svm = SVC()

svm.fit(X_train, y_train)
```

Fig 3 SVC

Random Forest - Random Forest is an ensemble learning system that trains several decision trees and outputs the class mode from each tree. By pooling predictions from several decision trees, it can detect phishing websites based on various factors.



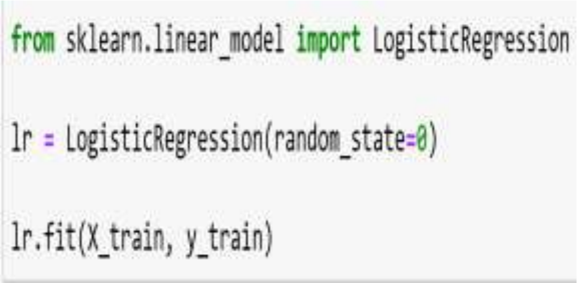
```
from sklearn.ensemble import RandomForestClassifier

rf = RandomForestClassifier(max_depth=2, random_state=0)

rf.fit(X_train, y_train)
```

Fig 4 Random forest

Logistic Regression - Although named logistic regression, it classifies binary data. Predicts the class likelihood of an input point. Based on particular parameters, logistic regression models may detect phishing websites.



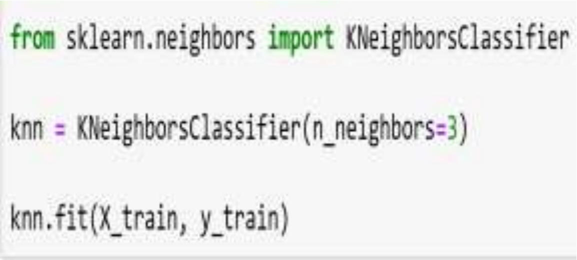
```
from sklearn.linear_model import LogisticRegression

lr = LogisticRegression(random_state=0)

lr.fit(X_train, y_train)
```

Fig 5 Logistic regression

K-Nearest Neighbors (KNN) - KNN is an easy classification method. The majority class of a data point's k-nearest neighbors in feature space determines its classification. KNN may use extracted characteristics to compare a website to known phishing or legal websites for phishing detection.



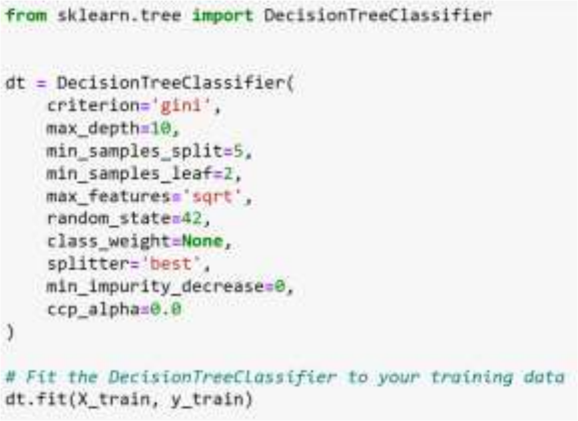
```
from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(n_neighbors=3)

knn.fit(X_train, y_train)
```

Fig 6 KNN

Decision Tree - A decision tree is a flowchart with internal nodes representing features, branches representing decisions depending on those features, and leaf nodes representing class labels. Decision trees may distinguish websites as phishing or legitimate depending on URL structure or content.



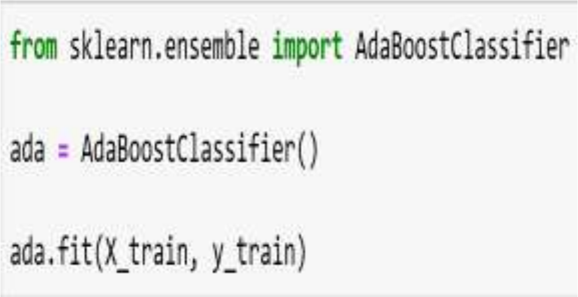
```
from sklearn.tree import DecisionTreeClassifier


dt = DecisionTreeClassifier(
    criterion='gini',
    max_depth=10,
    min_samples_split=5,
    min_samples_leaf=2,
    max_features='sqrt',
    random_state=42,
    class_weight=None,
    splitter='best',
    min_impurity_decrease=0,
    ccp_alpha=0.0
)

# Fit the DecisionTreeClassifier to your training data
dt.fit(X_train, y_train)
```

Fig 7 Decision tree

Adaboost - Ensemble learning approach Adaboost (Adaptive Boosting) builds a strong classifier from numerous weak classifiers. It improves suspicious website categorization for phishing detection by focusing on hard-to-classify occurrences.



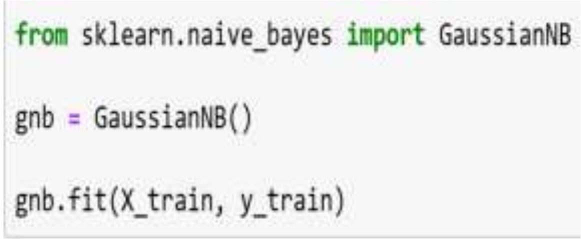
```
from sklearn.ensemble import AdaBoostClassifier

ada = AdaBoostClassifier()

ada.fit(X_train, y_train)
```

Fig 8 Adaboost

Naive Bayes - Naive Bayes is a Bayes' theorem-based probabilistic classification technique that assumes feature independence. It is effective and widely used for text categorization. Based on retrieved characteristics, Naive Bayes may detect phishing websites.



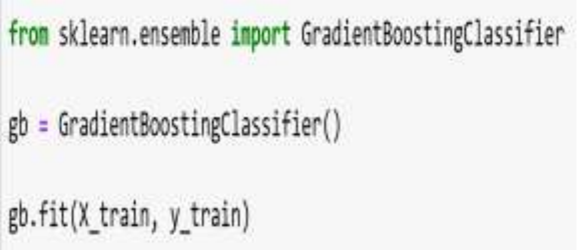
```
from sklearn.naive_bayes import GaussianNB

gnb = GaussianNB()

gnb.fit(X_train, y_train)
```

Fig 9 Naïve bayes

Gradient Boosting - Gradient Boosting gradually adds weak models to create a strong model. Goal is loss function minimization. Gradient boosting can combine weak learners to increase phishing detection accuracy.



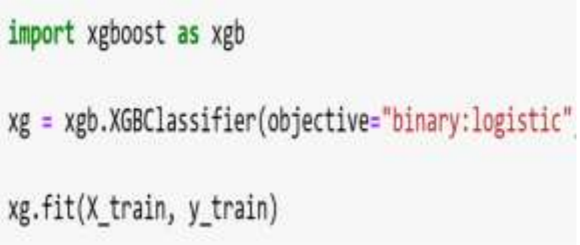
```
from sklearn.ensemble import GradientBoostingClassifier

gb = GradientBoostingClassifier()

gb.fit(X_train, y_train)
```

Fig 10 Gradient boosting

XGBoost - Gradient boosting is efficient and scalable using XGBoost. It's fast and good at machine learning competitions. Building an ensemble of weak models using XGBoost improves phishing detection.



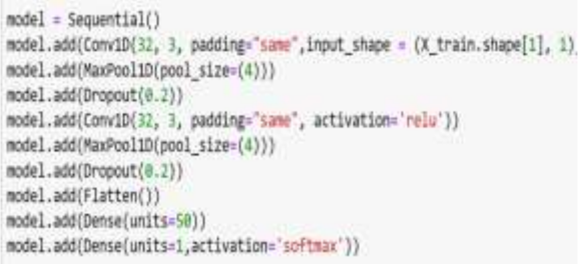
```
import xgboost as xgb

xg = xgb.XGBClassifier(objective="binary:logistic"

xg.fit(X_train, y_train)
```

Fig 11 XGboost

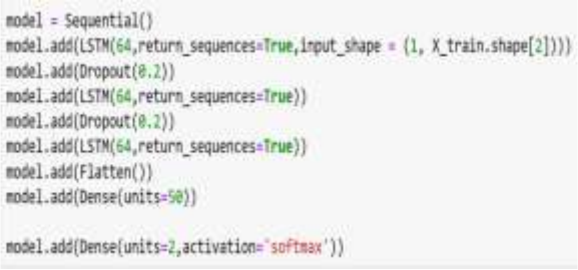
Convolutional Neural Network (CNN) - Deep learning CNN is used for image and pattern identification. CNN may detect phishing tendencies in webpage text or visuals.



```
model = Sequential()
model.add(Conv1D(32, 3, padding="same",input_shape = (X_train.shape[1], 1)
model.add(MaxPool1D(pool_size=(4)))
model.add(Dropout(0.2))
model.add(Conv1D(32, 3, padding="same", activation='relu'))
model.add(MaxPool1D(pool_size=(4)))
model.add(Dropout(0.2))
model.add(Flatten())
model.add(Dense(units=50))
model.add(Dense(units=1,activation='softmax'))
```

Fig 12 CNN

Long Short-Term Memory (LSTM) - LSTM (recurrent neural network) is ideal for sequential data processing. LSTM may predict phishing by analyzing URL or site content sequentially.

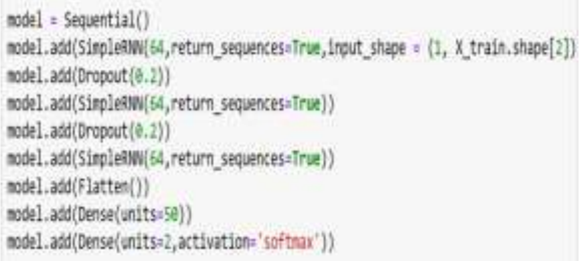


```
model = Sequential()
model.add(LSTM(64,return_sequences=True,input_shape = (1, X_train.shape[2])))
model.add(Dropout(0.2))
model.add(LSTM(64,return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(64,return_sequences=True))
model.add(Flatten())
model.add(Dense(units=50))

model.add(Dense(units=2,activation='softmax'))
```

Fig 13 LSTM

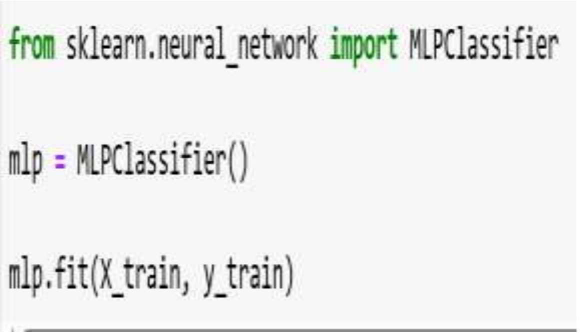
Deep Neural Network (DNN) - Multiple layers separate the input and output layers of a DNN. It captures complicated input data patterns and characteristics. Phishing detection may utilize DNN to classify features.



```
model = Sequential()
model.add(SimpleRNN(64,return_sequences=True,input_shape = (1, X_train.shape[2])
model.add(Dropout(0.2))
model.add(SimpleRNN(64,return_sequences=True))
model.add(Dropout(0.2))
model.add(SimpleRNN(64,return_sequences=True))
model.add(Flatten())
model.add(Dense(units=50))
model.add(Dense(units=2,activation='softmax'))
```

Fig 14 DNN

Multi-Layer Perceptron (MLP) - MLPs are feedforward neural networks. Multiple layers of linked nodes enable complicated learning. MLP may categorize phishing using extracted characteristics.



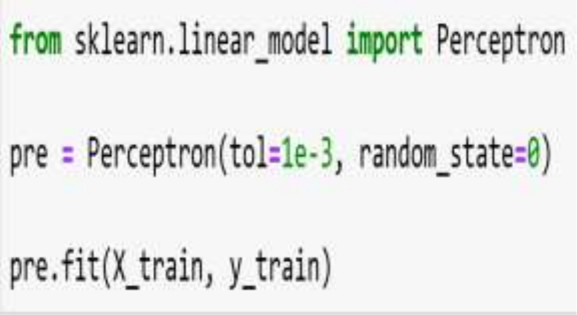
```
from sklearn.neural_network import MLPClassifier

mlp = MLPClassifier()

mlp.fit(X_train, y_train)
```

Fig 15 MLP

Perceptron - Simple artificial neural networks like perceptrons form the basis for more complicated models. Basic phishing detection may be done using this linear binary classification model.



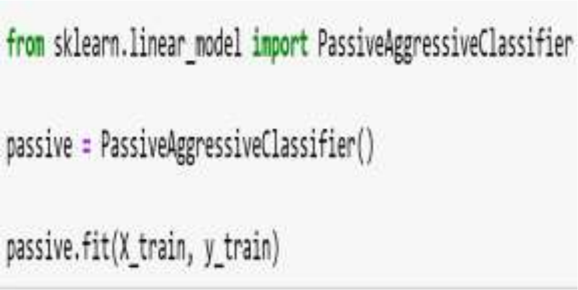
```
from sklearn.linear_model import Perceptron

pre = Perceptron(tol=1e-3, random_state=0)

pre.fit(X_train, y_train)
```

Fig 16 Perception

Passive Aggressive - Online learning algorithms called Passive Aggressive are utilized for categorization. They are important in phishing detection settings where data streams and models must adapt to shifting patterns.



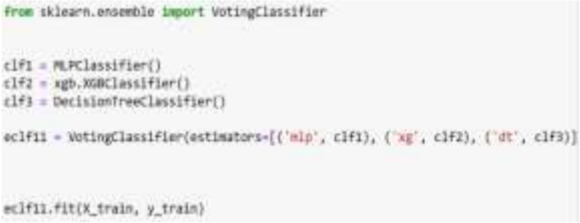
```
from sklearn.linear_model import PassiveAggressiveClassifier

passive = PassiveAggressiveClassifier()

passive.fit(X_train, y_train)
```

Fig 17 Passive aggressive

Voting Classifier - The Ensemble Voting Classifier predicts the class label by majority voting from many base estimators (e.g., models). Diverse models can improve phishing detection classification accuracy.



```
from sklearn.ensemble import VotingClassifier

clf1 = MLPClassifier()
clf2 = xgb.XGBClassifier()
clf3 = DecisionTreeClassifier()

eclf11 = VotingClassifier(estimators=[('mlp', clf1), ('xg', clf2), ('dt', clf3)]


eclf11.fit(X_train, y_train)
```

Fig 18 Voting 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)

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

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.

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

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:

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

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

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.

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

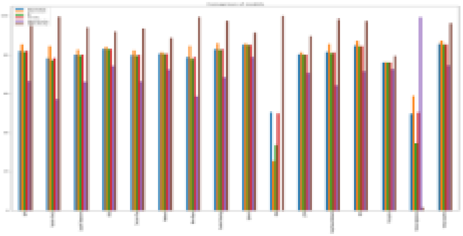
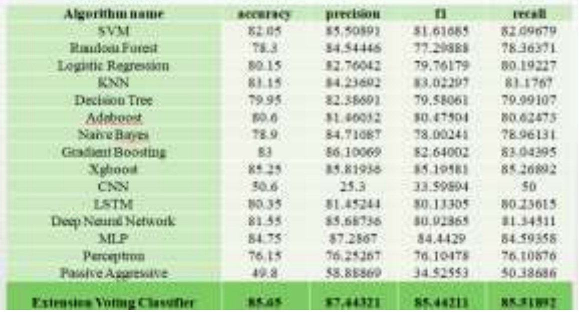


Fig 19 Performance graph

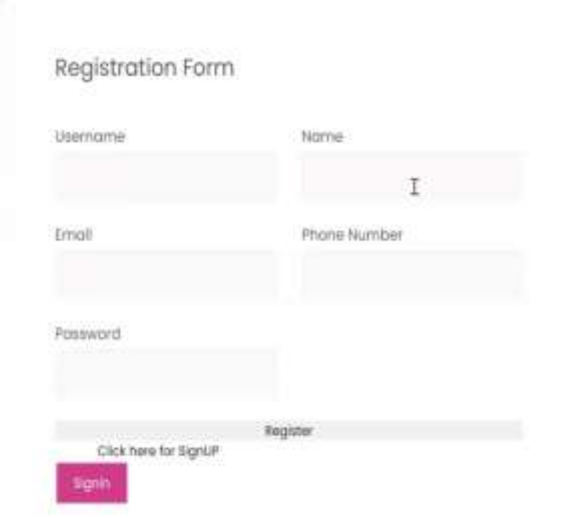


Algorithm name	accuracy	precision	f1	recall
SVM	82.05	85.50891	81.61685	82.09679
Random Forest	78.3	84.54446	77.29888	78.36371
Logistic Regression	80.15	82.76042	79.76179	80.19227
KNN	83.15	84.23692	83.02297	83.1767
Decision Tree	79.95	82.38691	79.58061	79.99107
Adaboost	80.6	81.46032	80.47504	80.62473
Naive Bayes	78.9	84.71087	78.00241	78.96131
Gradient Boosting	83	86.10069	82.64002	83.04395
Xgboost	85.25	85.81936	85.19581	85.26892
CNN	50.6	25.3	33.59894	50
LSTM	80.35	81.45244	80.13305	80.23615
Deep Neural Network	81.55	85.68736	80.92865	81.34511
MLP	84.75	87.2867	84.4429	84.59358
Perceptron	76.15	76.25267	76.10478	76.10876
Passive Aggressive	49.8	58.88869	34.52553	50.38686
Extension Voting Classifier	**85.65**	**87.44321**	**85.44211**	**85.51892**

Fig 20 Performance Evaluation

PHISHING OR NOT PHISHING? A SURVEY ON THE DETECTION OF PHISHING WEBSITES

Fig 21 Home page



Registration Form

Username

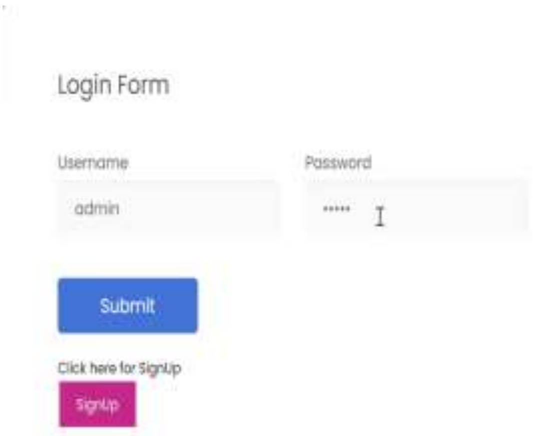
Name

Email

Phone Number

Password

Fig 22 Signin page



Login Form

Username: admin

Password: *****

Fig 23 Login page

Phishing Website Prediction

Type your url here

[illegible]

Fig 24 User input

You are safe.

Fig 25 Predict result for given input

5. CONCLUSION

The research studies phishing website detection technologies and methods. The goal is to comprehend the phishing detection landscape. The project shows the variety of phishing detection methods [40, 41, 42]. It also discusses assessment datasets, revealing the field's empirical basis. Furthermore, it highlights phishing detection research gaps that require further study. The project stresses the importance of feature selection based on strengths and shortcomings. It favors characteristics with strong discrimination power and attacker methods. An extension of the project, the Voting Classifier, detects phishing websites with 85.65% accuracy. The ensemble approach uses many classifiers to improve phishing detection and generalization across numerous contexts. An easy-to-use Flask interface with secure authentication improves system testing. This makes data entry and phishing detection system evaluation easy. The Flask interface streamlines testing and user interaction, making system assessment faster and easier.[59]

6. FUTURE SCOPE

The initiative promotes phishing detection study into its issues and shortcomings. This requires a thorough investigation of existing restrictions to build better countermeasures. The study focuses phishing defenses that include education since people are vulnerable. User education to detect and respond to phishing attempts is key to prevention. The initiative emphasizes the necessity to update and expand phishing website lists to improve detection [32, 34, 37]. This constant evolution keeps phishing detection methods current. The initiative encourages the development of novel phishing methods due to their dynamic nature. This proactive technique anticipates and detects new phishing methods. The initiative encourages phishing detection systems to use AI and machine learning. These technologies enable more advanced and adaptive detection. Existing detection algorithms will be evaluated and improved using larger and more diverse datasets. This method tests and validates detection systems under various real-world settings, improving their efficacy and applicability.

REFERENCES

[1] P. Zhang, A. Oest, H. Cho, Z. Sun, R. Johnson, B. Wardman, S. Sarker, A. Kapravelos, T. Bao, R. Wang, Y. Shoshitaishvili, A. Doupé, and G.-J. Ahn, ‘‘CrawlPhish: Large-scale analysis of client-side cloaking techniques in phishing,’’ in Proc. IEEE Symp. Secur. Privacy (SP), May 2021, pp. 1109–1124.

[2] ENISA. (2021). Cybersecurity for SMEs—Challenges and Recommendations. [Online]. Available: https://www.enisa.europa.eu/publications/ enisa-report-cybersecurity-for-smes

[3] Cisco. (2021). Cyber Security Threat Trends: Phishing, Crypto Top the List. [Online]. Available: https://umbrella.cisco.com/info/2021-cybersecurity-threat-trends-phishing-crypto-top-the-list

[4] M. Alsharnouby, F. Alaca, and S. Chiasson, ‘‘Why phishing still works: User strategies for combating phishing attacks,’’ Int. J. Hum.-Comput. Stud., vol. 82, pp. 69–82, Oct. 2015.

[5] Anti-Phishing Working Group—APWG. (2022). Phishing Activity Trends Report-1Q. [Online]. Available: https://docs.apwg.org/reports/ apwg_trends_report_q1_2022.pdf

[6] T. Berners-Lee, R. Fielding, and L. Masinter. Uniform Resource Identifier (URI): Generic Syntax, RFC 3986, Jan. 2005. [Online]. Available: https://www.rfc-editor.org/rfc/rfc3986.txt

[7] K. L. Chiew, K. S. C. Yong, and C. L. Tan, ‘‘A survey of phishing attacks: Their types, vectors and technical approaches,’’ Exp. Syst. Appl., vol. 106, pp. 1–20, Sep. 2018.

[8] B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, ‘‘Defending against phishing attacks: Taxonomy of methods, current issues and future directions,’’ Telecommun. Syst., vol. 67, no. 2, pp. 247–267, 2018.

[9] I. Qabajeh, F. Thabtah, and F. Chiclana, ‘‘A recent review of conventional vs. automated cybersecurity anti-phishing techniques,’’ Comput. Sci. Rev., vol. 29, pp. 44–55, Aug. 2018.

[10] E. Benavides, W. Fuertes, S. Sanchez, and M. Sanchez, ‘‘Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review,’’ in Developments and Advances in Defense and Security (Smart Innovation, Systems and Technologies), vol. 152, A. Rocha and R. P. Pereira, Eds. Berlin, Germany: Springer, 2020, pp. 51–64.

[11] D. Sahoo, C. Liu, and S. C. H. Hoi, ‘‘Malicious URL detection using machine learning: A survey,’’ 2017, arXiv:1701.07179.

[12] C. M. R. Da Silva, E. L. Feitosa, and V. C. Garcia, ‘‘Heuristic-based strategy for phishing prediction: A survey of URL-based approach,’’ Comput. Secur., vol. 88, Jan. 2020, Art. no. 101613.

[13] G. Varshney, M. Misra, and P. K. Atrey, ‘‘A survey and classification of web phishing detection

schemes,'' Secur. Commun. Netw., vol. 9, no. 18, pp. 6266–6284, Dec. 2016.

[14] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, ''A comprehensive survey of AI-enabled phishing attacks detection techniques,'' Telecommun. Syst., vol. 76, no. 1, pp. 139–154, Jan. 2021.

[15] A. Das, S. Baki, A. El Aassal, R. Verma, and A. Dunbar, ''SoK: A comprehensive reexamination of phishing research from the security perspective,'' IEEE Commun. Surveys Tuts., vol. 22, no. 1, pp. 671–708, 1st Quart., 2020.

[16] Z. Dou, I. Khalil, A. Khreishah, A. Al-Fuqaha, and M. Guizani, ''Systematization of knowledge (SoK): A systematic review of software-based web phishing detection,'' IEEE Commun. Surveys Tuts., vol. 19, no. 4, pp. 2797–2819, 4th Quart., 2017.

[17] A. K. Jain and B. B. Gupta, ''Phishing detection: Analysis of visual similarity based approaches,'' Secur. Commun. Netw., vol. 2017, pp. 1–20, Jan. 2017.

[18] 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. [19] Google Safe Browsing. Accessed: Oct. 10, 2022. [Online]. Available: https://safebrowsing.google.com/

[20] S. Bell and P. Komisarczuk, ''An analysis of phishing blacklists: Google Safe Browsing, OpenPhish, and PhishTank,'' in Proc. Australas. Comput. Sci. Week Multiconf., Feb. 2020, pp. 1–11.

[21] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, ''An empirical analysis of phishing blacklists,'' in Proc. 6th Conf. Email AntiSpam (CEAS), 2009, pp. 1–10.

[22] A. Oest, Y. Safaei, A. Doupé, G.-J. Ahn, B. Wardman, and K. Tyers, ''PhishFarm: A scalable framework for measuring the effectiveness of evasion techniques against browser phishing blacklists,'' in Proc. IEEE Symp. Secur. Privacy (SP), May 2019, pp. 1344–1361.

[23] A. Oest, Y. Safaei, P. Zhang, B. Wardman, K. Tyers, Y. Shoshitaishvili, and A. Doupé, ''PhishTime: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists,'' in Proc. 29th USENIX Secur. Symp., 2020, pp. 379–396.

[24] N. A. Azeez, S. Misra, I. A. Margaret, L. Fernandez-Sanz, and S. M. Abdulhamid, ''Adopting automated whitelist approach for detecting phishing attacks,'' Comput. Secur., vol. 108, Sep. 2021, Art. no. 102328.

[25] Y. Cao, W. Han, and Y. Le, ''Anti-phishing based on automated individual white-list,'' in Proc. 4th ACM Workshop Digit. Identity Manag., Oct. 2008, pp. 51–60.

[26] W. Han, Y. Cao, E. Bertino, and J. Yong, ''Using automated individual white-list to protect web digital identities,'' Exp. Syst. Appl., vol. 39, no. 15, pp. 11861–11869, Nov. 2012.

[27] A. K. Jain and B. B. Gupta, ''A novel approach to protect against phishing attacks at client side using auto-updated white-list,'' EURASIP J. Inf. Secur., vol. 2016, no. 1, pp. 1–11, Dec. 2016.

[28] L.-H. Lee, K.-C. Lee, H.-H. Chen, and Y.-H. Tseng, ''POSTER: Proactive blacklist update for anti-phishing,'' in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., Nov. 2014, pp. 1448–1450.

[29] 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.

[30] R. S. Rao and A. R. Pais, ''An enhanced blacklist method to detect phishing webwebwebsites,'' in Information Systems Security (Lecture Notes in Computer Science), vol. 10717, R. K. Shyamasundar, V. Singh, and J. Vaidya, Eds. Berlin, Germany: Springer, 2017, pp. 323–333.

[31] M. Sharifi and S. H. Siadati, ''A phishing webwebsites blacklist generator,'' in Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl., Mar. 2008, pp. 840–843.

[32] C. Whittaker, B. Ryner, and M. Nazif, ''Large-scale automatic classification of phishing pages,'' in Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS), 2010, pp. 1–14.

[33] G. Xiang, B. A. Pendleton, J. Hong, and C. P. Rose, ''A hierarchical adaptive probabilistic approach for zero hour phish detection,'' in Computer Security—ESORICS (Lecture Notes in Computer Science), vol. 6345, D. Gritzalis, B. Preneel, and M. Theoharidou, Eds. Berlin, Germany: Springer, 2010, pp. 268–285.

[34] 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, 2020.

[35] PhishTank. Accessed: Nov. 4, 2022. [Online]. Available: https://www.phishtank.org

[36] Curlie. Accessed: Nov. 4, 2022. [Online]. Available: https://curlie.org/

[37] S. Afroz and R. Greenstadt, ''PhishZoo: Detecting phishing webwebwebsites by looking at them,'' in Proc. IEEE 5th Int. Conf. Semantic Comput., Sep. 2011, pp. 368–375.

[38] J.-L. Chen, Y.-W. Ma, and K.-L. Huang, ''Intelligent visual similaritybased phishing webwebwebsites detection,'' Symmetry, vol. 12, no. 10, Oct. 2020, Art. no. 1681.

[39] K. T. Chen, J. Y. Chen, C. R. Huang, and C. S. Chen, ''Fighting phishing with discriminative keypoint features,'' IEEE Internet Comput., vol. 13, no. 3, pp. 56–63, May 2009.

[40] T.-C. Chen, S. Dick, and J. Miller, ''Detecting visually similar web pages: Application to phishing detection,'' ACM Trans. Internet Technol., vol. 10, no. 2, pp. 1–38, May 2010.

[41] J. Chen and C. Guo, ''Online detection and prevention of phishing attacks,'' in Proc. 1st Int. Conf. Commun. Netw. China, Oct. 2006, pp. 1–7.

[42] K. L. Chiew, E. H. Chang, S. N. Sze, and W. K. Tiong, ''Utilisation of webwebsite logo for phishing detection,'' Comput. Secur., vol. 54, pp. 16–26, Oct. 2015.

[43] M. Dunlop, S. Groat, and D. Shelly, ‘‘GoldPhish: Using images for content-based phishing analysis,’’ in Proc. 5th Int. Conf. Internet Monitor. Protection, 2010, pp. 123–128.

[44] A. Y. Fu, L. Wenyin, and X. Deng, ‘‘Detecting phishing web pages with visual similarity assessment based on Earth mover's distance (EMD),’’ IEEE Trans. Dependable Secure Comput., vol. 3, no. 4, pp. 301–311, Oct. 2006.

[45] M. Hara, A. Yamada, and Y. Miyake, ‘‘Visual similarity-based phishing detection without victim website information,’’ in Proc. IEEE Symp. Comput. Intell. Cyber Secur., Mar. 2009, pp. 30–36.

[46] C.-Y. Huang, S.-P. Ma, W.-L. Yeh, C.-Y. Lin, and C.-T. Liu, ‘‘Mitigate web phishing using website signatures,’’ in Proc. TENCON IEEE Region Conf., Nov. 2010, pp. 803–808.

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

[48] I. F. Lam, W. C. Xiao, S. C. Wang, and K. T. Chen, ‘‘Counteracting phishing page polymorphism: An image layout analysis approach,’’ in Advances in Information Security and Assurance (Lecture Notes in Computer Science), vol. 5576, J. H. Park, H. H. Chen, M. Atiquzzaman, C. Lee, T. Kim, and S. S. Yeo, Eds. Berlin, Germany: Springer, 2009, pp. 270–279.

[49] Y. Lin, R. Liu, D. M. Divakaran, J. Ng, Q. Chan, Y. Lu, Y. Si, F. Zhang, and J. Dong, ‘‘Phishpedia: A hybrid deep learning based approach to visually identify phishing webpages,’’ in Proc. 30th USENIX Secur. Symp., 2021, pp. 3793–3810.

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

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

[52] 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.

[53] 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.

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

[55] 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.

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

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

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

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

[60]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