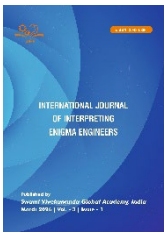
International Journal of Interpreting Enigma Engineers (IJIEE)
https://ejournal.svgacademy.org/index.php/ijiee/
https://doi.org/10.62674/ijiee.2025.v3i01.005
eISSN: 3048-6300
OPEN ACCESS

AN INTELLIGENT AI-DRIVEN FRAMEWORK FOR PHISHING DOMAIN DETECTION



Nithya Lakshmi N*, V Chaitanya

Original Article

[1,2]M V S R Engineering College, Hyderabad-501510, India

*Corresponding Author's Email: nithya_it@mvsrec.edu.in, 245120737050@mvsrec.edu.in

Abstract

The rapid increase in phishing attacks emerged as a major threat to users worldwide, with attackers utilizing different ways to mimic authentic domains and compromise user credentials. This paper recommends to develop an intelligent system that uses machine learning techniques to detect phishing domains particularly focusing on newly registered websites sourced from open and publicly available databases. By making use of the power of artificial intelligence, the proposed system assigns probability scores to identify the closeness of a domain to a genuine one. Moreover, it prioritizes the timely detection of new phishing domains improving the overall cyber security. This work deals with the significant requirement for an automated, intelligent tool to combat phishing attacks by proactively identifying malicious domains, ultimately safeguarding user credentials and cyber security measures.

Index Terms – Phishing, Artificial Intelligence, Machine Learning, Cyber security

Introduction

Phishing is an unreliable method of stealing customer personal as well as financial credentials using social and technological tricks. Generally attackers pretend to be trusted organizations by sending spoofed emails, messages, or notifications redirecting victims to false websites designed to capture sensitive information of them. Additionally, there is a possibility of installing the malicious software and deploying them on computers to compromise user systems and capture the login credentials associated with online accounts. In order to steal the information from the victims, phishers make use of an extensive range of attacking techniques like sending email messages, deceptive URLs, messages that are sent instantly, phone calls and text messages. The phishers prepare the structure of phishing content in such a way that is analogous to the actual content and swindle users in accessing those content so that they can obtain the sensitive data of the users. Thus the major goal of the phishing attacks is gaining certain personal information from the users either for the financial gain or for the use of identity theft that causes severe economic damage all over the world. The recent studies indicates that the sectors which are frequently targeted and affected are the financial institutions, payment services and webmail platforms.

Generally the cyber criminals obtain the sensitive and confidential information by generating an unauthorized replicas of genuine websites and emails that impersonates financial institutions or other organizations. They design the fraudulent e-mails by imitating the actual company's logos and slogans that seems to emerge authentic. The adaptable and flexible nature of HTML design makes it easy for the attackers to duplicate the images of the actual websites. This flexibility plays an important role for attackers in creating the opportunities that exploits the trusted brands, trademarks, and other corporate identifiers. To trap the users, Phisher sends these spooled emails in large volume to as many people as possible increasing the likelihood of deceiving unsuspecting users.

Traditional phishing detection methods mostly depend on human judgment, heuristic rules, or blacklist-based mechanisms however these methods struggle to keep pace with the increasingly sophisticated and rapidly evolving nature of phishing attacks. To overcome these limitations, machine learning based approaches have emerged as a robust solution enabling automated and precise detection of phishing attacks by identifying complex patterns and dynamically adapting to new attack strategies. The main objective of this paper is to build and train a machine learning model that can detect phishing websites using machine learning algorithms that are trained on a dataset of both legitimate and fraudulent websites. This directs to develop an effective phishing detection systems which automatically classify and warn users about potentially dangerous websites.

Literature Survey

In [1], the authors presented a machine learning based approach for detecting phishing websites using multiple classification algorithms for detecting malicious URLs. The performance of system was better rather than other ML methods but it lacks in handling layer volume of data.

The authors in [2] detected the phishing websites based on blacklisted dataset with lexical feature approach. But this reduces the performance with real time URLs.

In [3], the authors used the random forest classifier model using heuristic features and image analysis where the dataset is limited. Here in this paper the legitimate dataset is constructed only from Alexa's top websites.

Methodology

Proposed System

The system proposed takes the input URL provided by the user or any other external source and does the phishing detection employing the learning model used as test URL to identify whether it is a phishing URL or not. After performing the detection, the system provides the detection result indicating whether the URL is safe or flagged as a phishing URL where the legitimacy of the URL would be displayed on the screen.

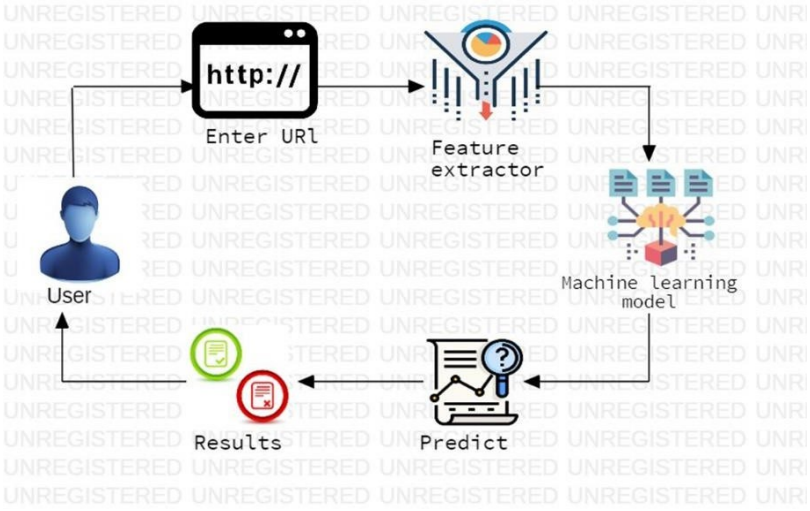


Figure 1: System Architecture

Machine Learning Algorithms

Logistic Regression**.** Logistic regression a commonly used classification algorithm when the class variable is of categorical nature. A supervised learning algorithm mainly helps in determining the relationship between features and the probability of a particular outcome. It utilizes the given set of independent variables and predicts the dependent variable which is categorical and provides the output as a probability value that ranges between 0 and 1.

Support Vector Machine**.** Support Vector Machine is employed for both classification & regression problems but often used in classification problems. SVM mainly aims to find a decision boundary line called hyperplane in N-dimension

space into classes. This helps to place the new data points in the correct space. The vectors which help to create the hyperplane are called as the support vectors.

Random Forest**.** Random Forest algorithm combines multiple classifiers to find a solution to a particular problem and used in both classification and regression problems. It combines many decision trees formed on different subsets of the dataset given and computes the average to make better predictions together. The algorithm predicts the final output considering the predictions made from each tree and based on the maximum voting of the predictions which makes Random Forest reliable and accurate for various tasks.

Naive Bayes Classifier**.** The Naive Bayes classifier grounded in the Bayesian framework empower it to navigate the intricate interplay of probabilities and attributes. Through a process of probabilistic assessment, this classifier appraises the likelihood of a given data point belonging to a specific class.

K Nearest Neighbor**.** K-Nearest neighbor algorithm considered to be a supervised machine learning algorithm performs both classification and regression problems. This algorithm predicts the values of new data points based on the similarity of features and assigns a value to the new data points considering how closely it matches with the data points in the training data set. This is basically done by creating an imaginary boundary line.

Ada boosting**.** It is an ensemble machine learning trick that assembles weak models (like basic decision trees) into a smart model for tasks like classifying things or predicting numbers. It learns from its mistakes and improves over time. Gradient Boosting**.** It is a boosting technique where weak models are combined step by step. The first model predicts the average, and each subsequent model corrects the errors of the previous one. The process continues until the errors become minimal. It requires numerical/categorical data and a differentiable loss function for updating predictions.

Result

In this model we have used different classification algorithms (Naive Bayes, Random Forest, KNN, xgboost, logistic regression, decision tree) as machine learning mechanism of the proposed systems. And in algorithms the gradient boosting algorithm classified the URLs with the highest accuracy of 97.4%. We have chosen a dataset with 11054 training instances and 31 features where 30 are independent features and 1 feature is dependent. Performances of the features and also algorithms compared. In the features out of the 30, HTTPS played highest importance and then Anchor URL followed by website traffic, links script, domains and prefix, suffixes.

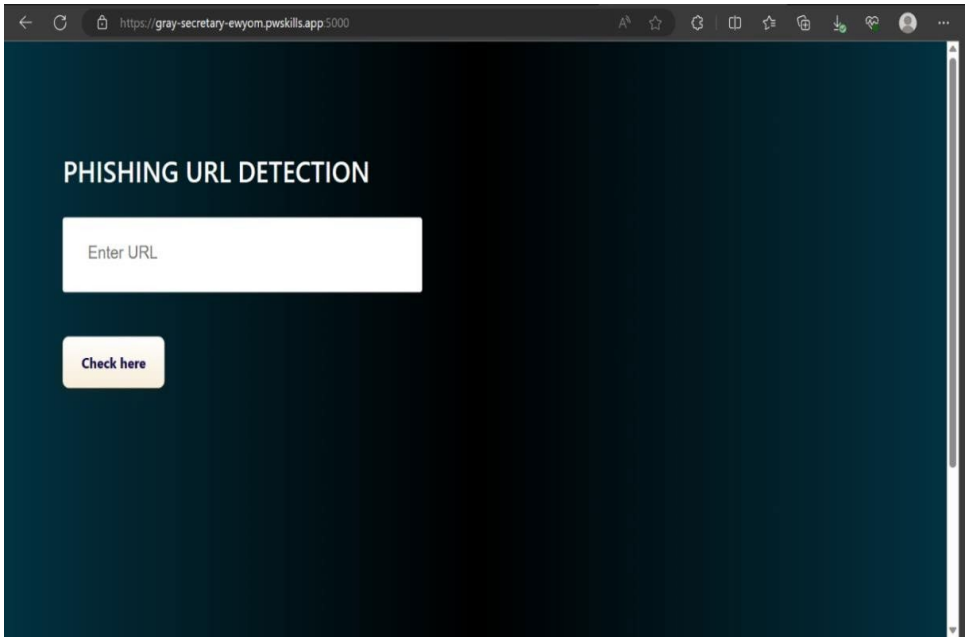


Figure 2. Home Page

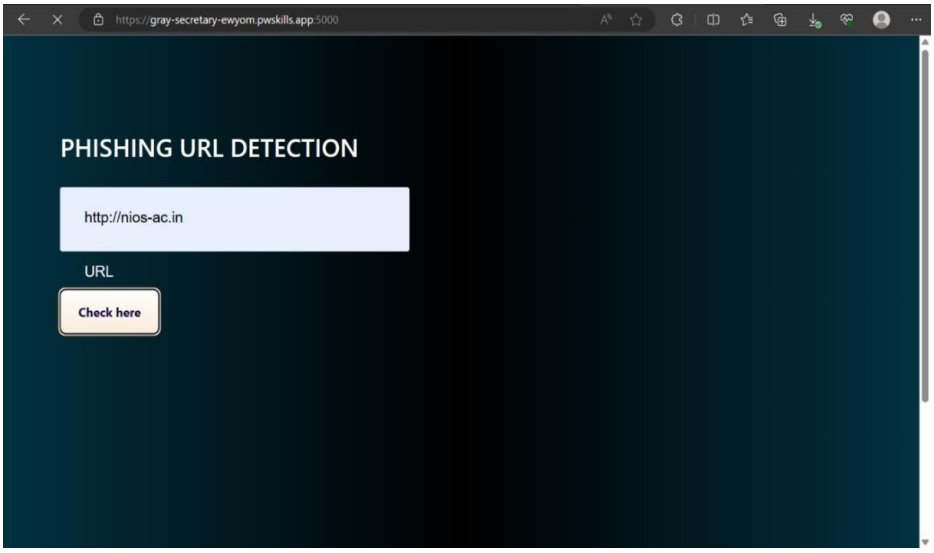


Figure 3. Give the URL as input

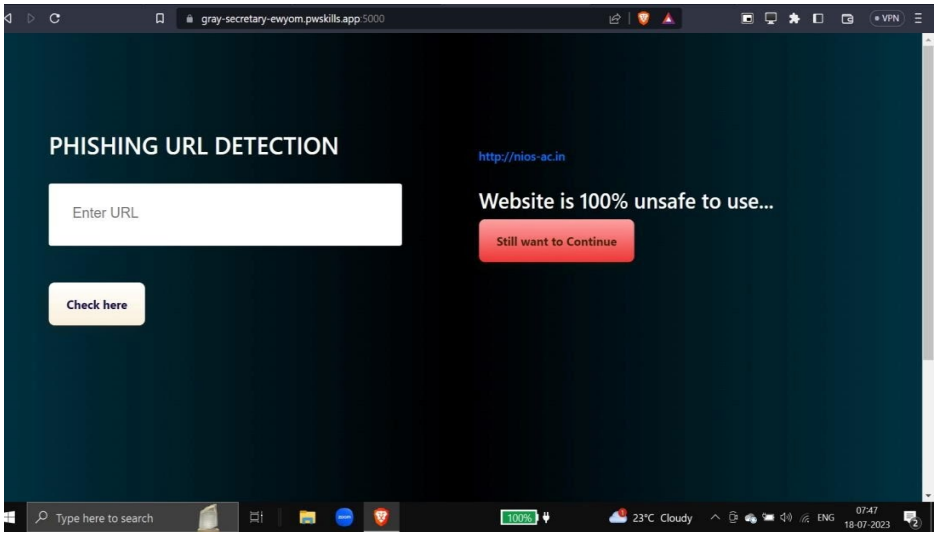


Figure 4. Result Page

	ML Model	Accuracy	f1_score	Recall	Precision
0	Gradient Boosting Classifier	0.974	0.977	0.994	0.986
1	CatBoost Classifier	0.972	0.975	0.994	0.989
2	Multi-layer Perceptron	0.971	0.974	0.992	0.985
3	XGBoost Classifier	0.969	0.973	0.993	0.984
4	Random Forest	0.967	0.970	0.992	0.991
5	Support Vector Machine	0.964	0.968	0.980	0.965
6	Decision Tree	0.961	0.965	0.991	0.993
7	K-Nearest Neighbors	0.956	0.961	0.991	0.989
8	Logistic Regression	0.934	0.941	0.943	0.927
9	Naive Bayes Classifier	0.605	0.454	0.292	0.997

Figure 5 Machine learning algorithms accuracy

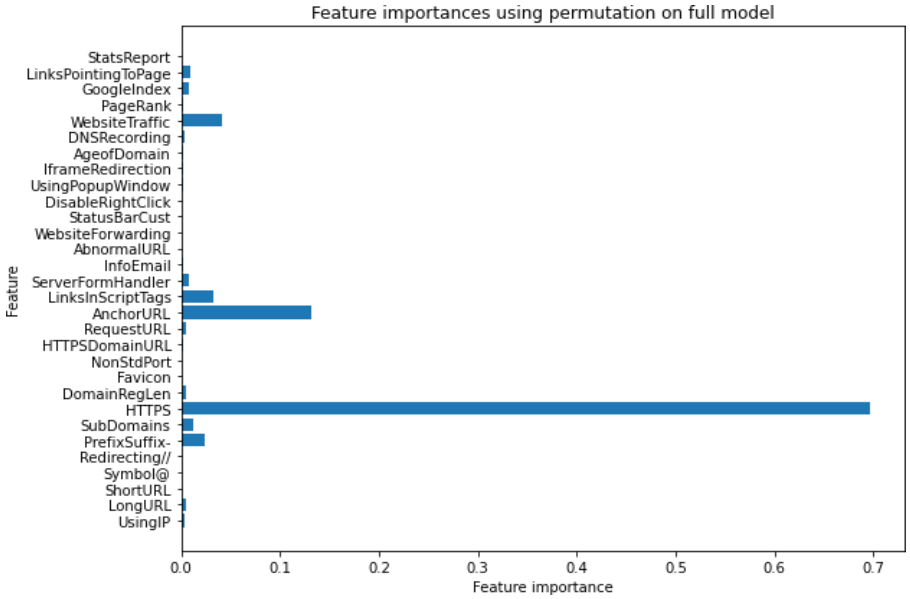


Figure 6 Feature importance using permutation on full model

Conclusion

Using the proposed system a phishing detection system is implemented where different learning algorithms were employed in the crucial task of constructing an efficient feature list to increase the accuracy of the detection system. The model used 31 features where 30 independent and 1 dependent variables were used enabling it to learn regarding the features affecting the URL legitimacy and also the performances of the model. Among the various machine learning models used, Gradient Boosting Classifier currently classify URL upto 97.4% respective classes.

References

1. Gandotra E., Gupta D, "An Efficient Approach for Phishing Detection using Machine Learning", Algorithms for Intelligent Systems, Springer, Singapore, 2021

2. Kumar, A. Santhanavijayan, B. Janet, B. Rajendran and B. S. Bindhumadhava, "Phishing Website Classification and Detection Using Machine Learning," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1–6, 10.1109/ICCCI48352.2020.9104161.

3. Rao RS, Pais AR. Jail-Phish: An improved search engine based phishing detection system. Computers & Security. 2019 Jun 1;83:246–67.

4. Jain A.K., Gupta B.B. "PHISH-SAFE: URL Features-Based Phishing Detection System Using Machine Learning", Cyber Security. Advances in Intelligent Systems and Computing, vol. 729, 2018.

5. Hung Le, Quang Pham, Doyen Sahoo, and Steven C.H. Hoi, "URLNet: Learning a URL Representation with Deep Learning for Malicious URL Detection", Conference'17, Washington, DC, USA, arXiv:1802.03162, July 2017