**Phishing Detection System Through Hybrid Machine Learning Based on URL**

Currently, numerous types of cybercrime are organized through the internet. Hence, this study mainly focuses on phishing attacks. Although phishing was first used in 1996, it has become the most severe and dangerous cybercrime on the internet. Phishing utilizes email distortion as its underlying mechanism for tricky correspondences, followed by mock sites, to obtain the required data from people in question. Different studies have presented their work on the precaution, identification, and knowledge of phishing attacks; however, there is currently no complete and proper solution for frustrating them. Therefore, machine learning plays a vital role in defending against cybercrimes involving phishing attacks. The proposed study is based on the phishing URL-based dataset extracted from the famous dataset repository, which consists of phishing and legitimate URL attributes collected from 11000+ website datasets in vector form. After preprocessing, many machine learning algorithms have been applied and designed to prevent phishing URLs and provide protection to the user. This study uses machine learning models such as decision tree (DT),linear regression (LR), random forest (RF), naive Bayes (NB), gradient boosting classifier (GBM), K-neighbors classifier (KNN), support vector classifier (SVC), and proposed hybrid LSD model, which is a combination of logistic regression, support vector machine, and decision tree (LR+SVC+DT) with soft and hard voting, to defend against phishing attacks with high accuracy and efficiency. The canopy feature selection technique with cross fold valoidation and GridSearchHyperparameterOptimizationtechniquesareusedwithproposed LSDmodel. Furthermore, to evaluate the proposed approach, different evaluation parameters were adopted, such as the precision, accuracy, recall, F1-score, and specificity, to illustrate the effects and efficiency of the models. The results of the comparative analyses demonstrate that the proposed approach outperforms the other models and achieves the best results.

Existing work

* This study uses a extension approach for phishing and URL detection and has an 85% accuracy rate; how ever, in recently, several automatic phishing detection mechanisms have been proposed
* . Existing method shortened URL features for the detection process, has 92% accuracy. Delta Phish is a phishing-detection mechanism. It uses several URL features to train supervised predictive models, and its accuracy rate is higher than 70%.
* Existing method proposes a Phish-Safe detection mechanism to detect malicious websites. This study used SVM and naive Bayes as a supervised-based machine learning approaches for phishing detection and achieved 90% accuracy

Proposed method

* Phishing URL-based cyberattack detection is proposed in this study to prevent crime and protect people’s privacy.
* • The dataset consists of 11000+ phishing URL attributes that help classify phishing URLs based on these attributes.
* • Machine learning models have been applied, such as decision tree (DT), linear regression (LR), naive Bayes (NB), random forest (RF), gradient boosting machine (GBM), support vector classifier (SVC), K-Neighbors classifier (KNN), and the proposed hybrid model (LR+SVC+DT) LSD with soft and hard voting, which can accurately classify the threats of phishing URLs.
* In proposed method we extent new method with cat boost classifier to improve accuracy.

**System requirement:**

### HARDWARE REQUIREMENTS:

* System : Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz
* Hard Disk : 1 TB.
* Input Devices : Keyboard, Mouse
* Ram : 4 GB.

### SOFTWARE REQUIREMENTS:

* Operating system : Windows XP/7/10.
* Coding Language : Python
* Tool : Anaconda
* Interface : flask webapp