



BVRIT HYDERABAD College Of Engineering for Women PHISHING WEBSITE DETECTOR

20WH1A0576 - RICHA CRISTINA

20WH1A0578 - SANDHYA M

20WH1A0579 - PHALASHI SINGH

20WH1A0581 - PUJITHA Y

20WH1A0582 - NEHARIKA T

01-04-2023



- Problem Statement:
- Binary classification model (phishing website or not) using Python Scikit-Learn that trains on the data and calculates the accuracy score on the test data. You have to use one or more of the classification algorithms to train a model on the phishing website dataset.



TECHSTACK

- Numpy
- Pandas
- Seaborn
- Matplotlib
- Sk-Learn
- Gradio
- Flask
- HTML/CSS/JS



ALGORITHMS USED

- Logistic Regression
- Support Vector Classification
- Random Forests
- Gradient Boosting
- KNN - K nearest neighbor
- Gaussian NB classification
- ADA Boost Classification
- Linear Discrimination Analysis
- Quadratic Discrimination Analysis
- Decision Tree classification



- **Random Forests:**
- It is used for both classification and regression tasks. The Random Forest algorithm combines the predictions of multiple decision trees, each of which is trained on a different subset of the features.
- The features that can be used to train the model include website content, URL structure, SSL certificates, and other factors that can be indicative of a phishing website

• Decision Trees:

- During the training process, the decision tree algorithm learns the optimal splits based on the information gain and Gini index criteria.
- Decision trees are a powerful machine learning algorithm that can be used to detect phishing websites by recursively splitting the data based on the values of the features.

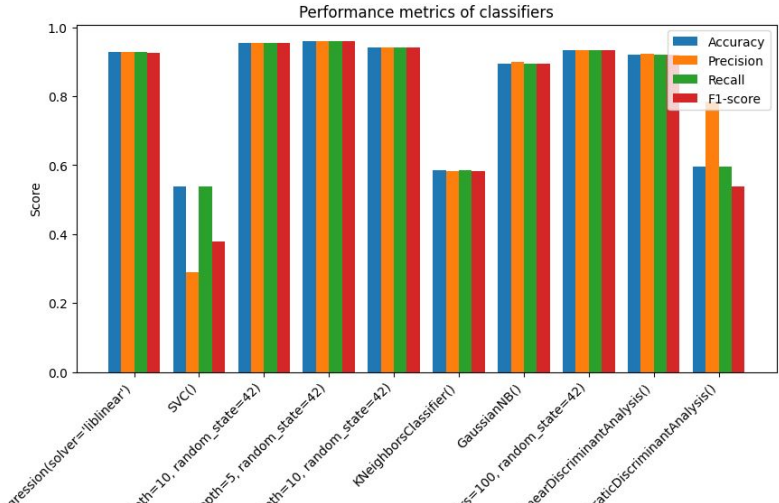


- **Gradient Boost:**
- It iteratively improves a weak learner (e.g., a decision tree) to create a strong learner.
- Gradient Boosting over other algorithms like Random Forests and Decision Trees is that it can handle imbalanced datasets.

Comparison Table:

	Classifier	accuracy	f1	recall	precision	
0	LogisticRegression(solver='liblinear')	93.198263	93.181711	93.198263	93.222582	
1	SVC()	56.005789	40.211948	56.005789	31.366484	
2	RandomForestClassifier(max_depth=10, random_st...	95.586107	95.575870	95.586107	95.619757	
3	GradientBoostingClassifier(max_depth=5, random...	96.020260	96.016925	96.020260	96.021870	
4	DecisionTreeClassifier(max_depth=10, random_st...	94.609262	94.609499	94.609262	94.609761	
5	KNeighborsClassifier()	59.659913	59.545274	59.659913	59.477716	
6	GaussianNB()	88.277858	88.320226	88.277858	89.046020	
7	AdaBoostClassifier(n_estimators=100, random_st...	93.704776	93.694339	93.704776	93.712793	
8	LinearDiscriminantAnalysis()	92.619392	92.599722	92.619392	92.647063	
9	QuadraticDiscriminantAnalysis()	56.874096	50.467050	56.874096	78.222135	

Graphical Representation:



OUTPUT

Phishing Website Detector

Enter a website URL to determine whether it's safe or phishing

url

<http://www.youtube.com>

Clear

Submit

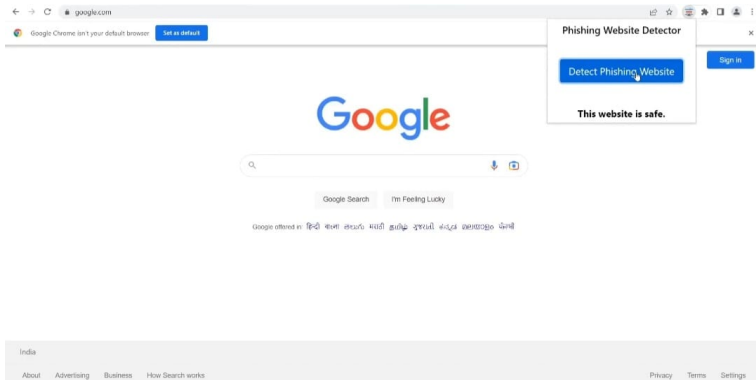
output

1

Flag



LEGITIMATE WEBSITE



← → ↻ google.com

Google Chrome isn't your default browser [Set as default](#)

Phishing Website Detector

[Detect Phishing Website](#)

[Sign in](#)

This website is safe.

Google

Google Search [I'm Feeling Lucky](#)

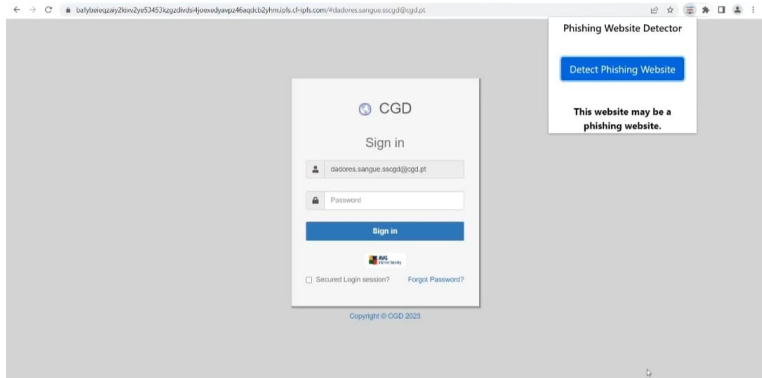
Google offered in [हिन्दी](#) [বাংলা](#) [தமிழ்](#) [മലയാളം](#) [ગુજરાતી](#) [ಕನ್ನಡ](#) [ಕೊಡಗ](#)

India

[About](#) [Advertising](#) [Business](#) [How Search works](#)

[Privacy](#) [Terms](#) [Settings](#)

PHISHING WEBSITE





THANK YOU