

PROJECT REPORT – WEB PHISING DETECTION

Team Id: PNT2022TMID20607

Batch Id: B1-1M3E

Team Members:

- VARSHINI RAMESH – TEAM LEADER
- SNEHA DHARSHINI U
- RUBINI V
- SNEGHA B
- THARANISKA S

Table of Content

- * Introduction
- * Installation
- * Directory Tree
- * Result
- * Conclusion

Introduction

The Internet has become an indispensable part of our life, However, It also has provided opportunities to anonymously perform malicious activities like Phishing. Phishers try to deceive their victims by social engineering or creating mockup websites to steal information such as account ID, username, password from individuals and organizations. Although many methods have been proposed to detect phishing websites, Phishers have evolved their methods to escape from these detection methods. One of the most successful methods for detecting these malicious activities is Machine Learning. This is because most Phishing attacks have some common characteristics which can be identified by machine learning methods. To see project click [\[here\]\("/"\)](#).

Installation

The Code is written in Python 3.6.10. If you don't have Python installed you can find it [\[here\]\(https://www.python.org/downloads/\)](https://www.python.org/downloads/). If you are using a lower version of Python you can upgrade using the pip package, ensuring you have the latest version of pip. To install the required packages and libraries, run this command in the project directory after [\[cloning\]\(https://www.howtogeek.com/451360/how-to-clone-a-github-repository/\)](https://www.howtogeek.com/451360/how-to-clone-a-github-repository/) the repository:

pip install -r requirements.txt

Directory Tree

...

```
|— pickle
|   |— model.pkl
|   — static
|   |— styles.css
|   — templates
|       |— index.html
|       — Phishing URL Detection.ipynb
|       — Procfile
|       — README.md
|       — app.py
|       — feature.py
|       — phishing.csv
|       — requirements.txt
```

...

Technologies Used

- NUMPY
- PANDAS
- MATPLOTLIB
- SCIKIT
- FLASK

Result

Accuracy of various model used for URL detection

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	XGBoost Classifier	0.969	0.973	0.993	0.984
3	Multi-layer Perceptron	0.969	0.973	0.995	0.981
4	Random Forest	0.967	0.971	0.993	0.990
5	Support Vector Machine	0.964	0.968	0.980	0.965
6	Decision Tree	0.960	0.964	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

Conclusion

1. The final take away from this project is to explore various machine learning models, perform Exploratory Data Analysis on phishing dataset and understanding their features.
2. Creating this notebook helped me to learn a lot about the features affecting the models to detect whether URL is safe or not, also I came to know how to tune model and how they affect the model performance.
3. The final conclusion on the Phishing dataset is that some features like "HTTPS", "AnchorURL", "WebsiteTraffic" have more importance to classify URL as phishing or not.
4. Gradient Boosting Classifier correctly classifies URL up to 97.4% across respective classes and hence reduces the chance of malicious attachments.