PHISHING WEBSITE URL DETECTION (CONTENT-BASED) USING STREAMLIT

INTRODUCTION

The rise of online phishing attacks poses a significant threat to internet users' security. Phishing websites mimic legitimate sites to steal sensitive information such as login credentials, financial data, etc. Detecting and managing these malicious URLs is crucial to safeguard users' privacy and prevent cybercrimes. Our project aims to develop a comprehensive Phishing Website URL Detection Management System using Streamlit, a user-friendly framework for building interactive web applications in Python.

OBJECTIVE

The primary objective of this project is to develop a user-friendly web application that utilizes machine learning algorithms to detect phishing website URLs.

CANDIDATES

TEAM HEAD:
110120104014 - BALAJI S

TEAM MEMBERS:
110120104004 - ABDUL QAYYUM
110120104032 - MOHAMED ZAFER
110120104035 - MOHAMMED IRFAN

SUPERVISOR

MS. U. SUBATHRA,
ASSISTANT PROFESSOR, CSE

KEY FEATURES

- URL Classification
- Real-Time Detection
- Database Management
- User Interface

TECHNOLOGIES USED

- Python
- Streamlit
- Machine Learning (Scikit-Learn)
- Web Scraping (bs4 - BeautifulSoup)
- SQLite

METHODOLOGY

1. Gather a dataset of URLs labelled as either a legitimate or a phishing.
2. Extract relevant features from the URLs such as length of the URL, domain age, presence of HTTPS etc,.
3. Train machine learning models such as Random Forests, SVMs, Decision Trees, etc., using the extracted features.
4. Develop a web application using Streamlit, incorporating modules for URL input, feature extraction, model prediction, and result display.
5. Test the application thoroughly to identify and fix any bugs or issues.

BENEFITS

- Strengthened online security by promptly identifying and blocking phishing URLs.
- Continuous improvement in detection accuracy through advanced machine learning techniques.

CONCLUSION

Our "Phishing Website URL Detection Management System using Streamlit" empowers users to combat online phishing threats effectively. With its intuitive interface and powerful detection capabilities, it serves as a valuable tool in ensuring online security and privacy.

FUTURE SCOPE

- Integration with browser extensions for real-time protection during web browsing.
- Implementation of more sophisticated algorithms for enhanced detection accuracy.
- Collaboration with cybersecurity organizations to expand the database of known phishing URLs.

PHISHING WEBSITE URL DETECTION (CONTENT-BASED) USING STREAMLIT

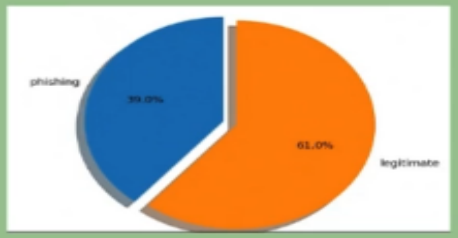
COMPARISON

ALGORITHM	ACCURACY	PRECISION	RECALL
GAUSSIAN NAIVE BAYES	0.85	0.87	0.82
K-NEAREST NEIGHBORS (K-NN)	0.92	0.91	0.94
NEURAL NETWORKS	0.95	0.94	0.96
SUPPORT VECTOR MACHINE	0.89	0.88	0.90
ADABOOST	0.93	0.92	0.94
DECISION TREE	0.87	0.85	0.89

MACHINE LEARNING MODELS

- Gaussian Naive Bayes
- K-Nearest Neighbors
- Neural Networks
- Support Vector Machine
- AdaBoost
- Decision Tree
- Gaussian Process

PIE CHART



Results

We used 7 different ML classifiers of scikit-learn and tested them implementing k-fold cross validation.Firstly obtained their confusion matrices, then calculated their accuracy, precision and recall scores.Comparison table is below:

	accuracy	precision	recall
NB	0.4985	0.0388	0.9137
SVM	0.9343	<NA>	0.1257
DT	0.9842	0.6715	0.7010
RF	0.9908	0.8779	0.6721
AB	0.9782	0.5284	0.1683
NN	0.9620	<NA>	0.1571
K-NN	0.9781	0.4807	0.1434

RESULTS

Phishing Website URL Detection using Machine Learning

This ML-based app is developed for educational purposes. Objective of the app is detecting phishing websites only using content data. Not URL! You can see the details of approach, dataset, and feature set if you click on *"See The Details"*.

PROJECT DETAILS

EXAMPLE PHISHING URLs: