PHISHING WEBSITE DETECTION by MACHINE LEARNING TECHNIQUES

ARTHI VENKATESAN

INTRODUCTION

- Phishing is one of the most prevalent social engineering and cyber attack techniques employed by attackers.
- In phishing attacks, malicious actors deceive unsuspecting online users into divulging sensitive information, which is then exploited for fraudulent purposes.
- To protect against phishing, users need to be aware of phishing websites and maintain a blacklist of known phishing sites.
- However, a more effective approach is to detect newly emerging phishing websites early through machine learning and deep neural network algorithms.
- Among these methods, machine learning-based techniques have proven to be the most effective for phishing detection.
- Despite the availability of these detection methods, online users continue to fall victim to phishing attacks, revealing sensitive information on fraudulent websites.

OBJECTIVE

- Phishing websites are a prevalent social engineering tactic that involves mimicking legitimate URLs and web pages to deceive users.
- The goal of this project is to develop machine learning models and deep neural networks capable of predicting phishing websites based on a carefully curated dataset.
- The dataset comprises both phishing and benign (legitimate) website URLs, collected for training purposes.
- Relevant features are extracted from the URLs and website content to serve as inputs for the machine learning models.
- The performance of each trained model is evaluated and compared using appropriate metrics.
- The objective is to identify the most accurate and effective approach for detecting phishing websites among the various models and techniques explored.

APPROACH

- Acquire a dataset containing both phishing and legitimate website URLs from open-source platforms.
- Develop code to extract relevant features from the URL database for model training.
- Perform exploratory data analysis (EDA) and preprocess the dataset using appropriate techniques.
- Split the dataset into training and testing subsets for model evaluation.
- Implement selected machine learning algorithms (e.g., Support Vector Machines, Random Forests) and deep neural network models (e.g., Autoencoders) on the training data.
- Write code to evaluate the trained models' performance using appropriate accuracy metrics.
- Compare the results obtained from the different models and identify the most effective approach for phishing website detection.
- Analyze the strengths and weaknesses of each model and provide recommendations based on the evaluation outcomes.

DATA SET COLLECTION

- Dataset has been collected from the website URL given in the project itself . The URL is :Phishing Websites Dataset - Mendeley Data
- This datasets consists on the categories on artificial intelligence, computer security, privacy.
- It has 2 major files in which the full variant consists of 88,657 instances of legitimate and phishing websites instances
- The small variant has 58,645 instances containing legitimate and phishing websites instances.
- Both has the total of 111 features.

FEATURE SELECTION

The following category of features are selected:

- Address Bar based Features
- Domain based Features
- HTML & Javascript based Feature
- Address Bar based Features considered are:
- Domain of URL
- Redirection ‘//’ in URL
- IP Address in URL
- ‘http/https’ in Domain name
- ‘@’ Symbol in URL
- Using URL Shortening Service
- Length of URL
- Prefix or Suffix "-" in Domain
- Depth of URL

FEATURE SELECTION(CONT.)

Domain based Features considered are:

• DNS Record • Age of Domain • Website Traffic • End Period of Domain •

• HTML and JavaScript based Features considered are:

Iframe Redirection • Disabling Right Click • Status Bar Customization • Website Forwarding

• All together 17 features are extracted from the dataset.

MACHINE LEARNING MODELS

• The task of detecting phishing websites falls under the category of supervised machine learning problems.

• Supervised learning encompasses two main types of problems: classification and regression.

• In this case, it is a classification problem, where each input URL needs to be classified as either phishing (labeled 1) or legitimate (labeled 0).

• Several machine learning classification models are employed to train on the dataset, including:

- Decision Tree classifier
- Random Forest classifier
- Multilayer Perceptron (Neural Network) classifier
- XGBoost (Extreme Gradient Boosting) classifier
- Autoencoder Neural Network model
- Support Vector Machines (SVM) classifier

• Each of these models will be trained on the labeled dataset, with the goal of accurately classifying new, unseen URLs as either phishing or legitimate.

MODEL EVALUATION

• The models are evaluated, and the considered metric is accuracy.

• Below Figure shows the training and test dataset accuracy by the respective models:

• For the above it is clear that the XGBoost model gives better performance. The model is saved for further usage.

NEXT STEPS

Working on this project is very knowledgeable and worth the effort.

• Through this project, one can know a lot about the phishing websites and how they are differentiated from legitimate ones.

• This project can be taken further by creating a browser extensions of developing a GUI.

• These should classify the inputted URL to legitimate or phishing with the use of the saved model.

THANK YOU