

Internal Faculty: Dr. Rajesh Dhumal

**Title: Detection of Phishing Websites using Machine Learning Techniques**

Introduction:

The rapid growth of the internet and online services has significantly enhanced our daily lives, enabling everything from shopping to banking and many more. However, with the evolution of the E-services, the Phishing attacks has also increased.

Phishing is a fraudulent attempt to obtain sensitive information such as usernames, passwords, financial details and then use that information for various threats like identity theft etc. These attacks can lead to severe financial loss and compromise personal data.

Traditional methods to combat phishing, such as List based methods or Heuristic-based detection, have limitations in terms of efficiency. With the evolution of new phishing techniques, these methods fall short in identifying the fraudulent websites. Hence there is a growing need for more advanced techniques for the detection of phishing websites.

Machine Learning provides a solution to this problem. By leveraging large datasets and algorithms, ML models can learn and recognize the patterns and features that distinguishes phishing websites from the legitimate ones. This project aims to explore and implement machine learning techniques to detect phishing websites, contributing to the efforts to secure digital space.

Problem Statement:

Phishing attacks remains one of the most prevalent and damaging forms of cyber threats. Despite existing measures, many phishing websites go undetected until significant harm is done. Traditional methods like List based methods or Heuristic methods or Visual Similarity methods, struggle to keep pace with the rapidly evolving tactics used by the cyber criminals. The primary challenge lies in developing a robust, adaptive and efficient detection system that can accurately identify phishing websites with minimal false positives.

Objectives:

1. Discover the best phishing website detection techniques to help the security manager easily select the top technique among the anti-phishing approaches for their security systems.

2. A good review paper is needed that focuses on the techniques, data sets, and algorithms used by the scholars in the regarded area.

Methodology:

**Dataset Collection and EDA:** A large collection of false and legitimate URLs are to be collected and perform necessary EDA and preprocessing to clean and prepare the data for further processing.

**Feature Extraction:** Identify and extract relevant features from websites such as Address Bar-based features, Domain based features and HTML based features which can help in distinguishing the phishing websites from the legitimate ones.

**Model development:** Develop and train model based on the preprocessed data using machine learning algorithms like Decision tree, random forest, SVM and others. Experiment with different algorithms to determine the most effective approach.

Expected Outcomes:

* Enhanced Understanding of Phishing Website detection techniques.
* Well defined feature set for phishing detection.
* Accurate machine learning models to detect phish websites based on identified features.
* New approach to the Detection of phishing websites through feature extraction.
* Optimized Model Performance

References:

<https://www.sciencedirect.com/science/article/pii/S1319157823000034>

<https://www.mdpi.com/2076-3417/13/8/4649>

<https://www.researchgate.net/publication/256841808_Phishing_Detection_A_Literature_Survey>

<https://link.springer.com/article/10.1007/s10115-022-01672-x>

<https://ijcrt.org/papers/IJCRTI020051.pdf>