**PHISHING WEBSITE DTECTION**

**OBJECTIVE:**

The objective of phishing website detection using machine learning is to develop an automated system that can accurately and efficiently identify websites that are designed to steal sensitive information from users. This application is to investigate a specific problem of whether it is valuable or not to use machine learning techniques to predict the type of website.

**ABSTRACT:**

Phishing websites are a serious threat to online security, as they attempt to steal sensitive information from unsuspecting users. To combat this threat, researchers have developed various techniques for detecting phishing websites, including machine learning algorithms. Machine learning algorithms can be trained on large datasets of phishing and legitimate websites to learn patterns and characteristics that distinguish between the two. These algorithms can then be used to identify and block phishing websites before users can be victimized. One approach to phishing website detection using machine learning involves feature extraction, where various features of a website such as URL structure, domain age, and content are analyzed to identify phishing websites. Another approach involves using deep learning algorithms to automatically extract features and learn complex patterns in website data. Overall, machine learning-based phishing website detection techniques have shown promising results, achieving high accuracy rates and outperforming traditional rule-based methods. With further research and development, these techniques have the potential to become an important tool in the fight against online phishing attacks.

Keywords: Machine learning techniques, Xgboost, Gradient boosting, Adaboost, SVM, Random forest, evaluation.

**EXISTING SYSTEM:**

Existing systems for phishing website detection using machine learning often rely on supervised learning techniques that require labeled data for training. This means that a large dataset of both legitimate and phishing websites must be manually labeled, which can be time-consuming and costly.

**DISADVANTAGES:**

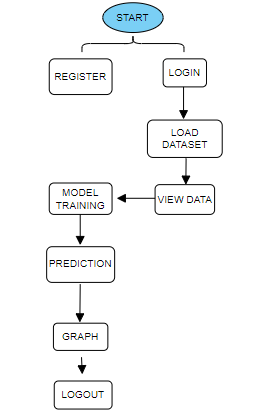
* Disadvantage is that these systems may be vulnerable to adversarial attacks, where attackers can manipulate the website features to evade detection. Additionally, machine learning-based systems may require significant computational resources, which could limit their scalability.
* Finally, these systems may produce false positives or false negatives, which can impact the user experience and reduce the effectiveness of the system. These limitations highlight the need for ongoing research and development in the field of machine learning-based phishing website detection.

**PROPOSED SYSTEM:**

The proposed system for phishing website detection using machine learning algorithms aims to overcome the limitations of existing systems. One approach is to use supervised learning techniques that do not require labelled data for training. This can reduce the time and cost of data labelling and improve scalability.

* Another approach is to incorporate multiple machine learning algorithms to enhance the accuracy and robustness of the system. The system can also be augmented with additional features such as website behaviour analysis and user behaviour monitoring to improve its ability to detect phishing websites.
* Overall, the proposed system aims to improve the accuracy, efficiency, and effectiveness of phishing website detection using machine learning algorithms.

**BLOCK DIAGRAM:**



**SYSTEM SPECIFICATIONS:**

# H/W SPECIFICATIONS:

# Processor : I3/Intel Processor

* RAM : 4GB (min)
* Hard Disk : 128 GB
* Key Board : Standard Windows Keyboard
* Mouse : Two or Three Button Mouse
* Monitor : Any

**S/W SPECIFICATIONS:**

* Operating System : Windows 7+
* Server-side Script : Python 3.6+
* IDE : PyCharm.
* Libraries Used : Pandas, Numpy, Matplotlib, OS.