**Classification and detecting Phishing website**

**Team Members:**

Latha Narayani Balaji

Mohammed Johny Shaik

Kushal Muktala

**Objectives:**

A phishing website is a social manipulation technique that imitates legitimate (URLs) and webpages. The purpose of this study is to use the dataset collected to train machine learning methods and deep artificial neural networks (Ann) to anticipate phishing websites. The required URL and website filters based on the content, attributes(features) are extracted from the dataset, which includes all phishing URLs of websites. Each model's performance is determined by comparison.

**Motivation:**

In 2020, the most common sort of cybercrime was phishing. - Federal Bureau of Investigation.

The best firewall or antivirus software can be implemented by any firm. However, most businesses do not have effective information security awareness initiatives for their employees. A single negligent employee can do a lot of damage to the organization.

**Significance:**

Phishing detection systems are essential in providing people to have a safe and secure online environment.

**Dataset:**

[https://archive.ics.uci.edu/ml/datasets/phishing+websites#](https://archive.ics.uci.edu/ml/datasets/phishing+websites)

**Classifiers Used:**

1. Random Forest Classifier
2. Decision tree
3. Logistic regression
4. KNN
5. XGBoost Classifier
6. ANN
7. Support Vector Machines
8. Multilayer perceptron

**Features:**

We are targeting close to 17 feature combing based on content and URL.

**GitHub URL:**

<https://github.com/Lathabalaji/NLP_Project>