SOLUTION

The project will follow the same approach as used in all ML project. We'll go through different stages of data collection, feature extraction, training and finally deployment of trained model.

* Data Collection
* Feature Extraction
* Model training & evaluation
* Deployment

Data Collection:

* Use open-source WHOIS databases and APIs to collect information about newly registered domain names. WHOIS databases provide details about domain registrants, registration dates, and more.
* For labeled data, gather examples of known phishing domains and legitimate domains. You can find datasets online that contain labeled phishing URLs. Websites like PhishTank provide such datasets

Feature Engineering:

Extract relevant features from the collected data.

These may include:

* Domain Features: Age of the domain, registrar information, domain length, and the presence of certain keywords.
* WHOIS Features: Registrant country, registration date, and expiry date

Model Training:

Classification problem comes under supervised machine learning. After feature extraction we'll train multiple ML models using our data and choose the model which gives us best accuracy.

The machine learning models considered to train the dataset in this projects are :

* Decision Tree
* Random Forest
* Multilayer Perceptron

Testing and Validation:

The trained model is tested and validated using a separate dataset to ensure its accuracy and effectiveness in detecting phishing domains.