

# **WEB PHISHING DETECTION**

## **PROJECT DEVELOPMENT PHASE**

### **SPRINT – 2**

#### **PRIOR KNOWLEDGE**

**One should have knowledge on the following Concepts:**

##### **Supervised and Unsupervised learning:**

Supervised machine learning relies on labeled input and output training data, whereas unsupervised learning processes unlabelled or raw data. In supervised machine learning the model learns the relationship between the labeled input and output data. In unsupervised learning, only input data is provided to the model.

##### **Regression Classification and Clustering:**

Regression and Classification are types of supervised learning algorithms while Clustering is a type of unsupervised algorithm. When the output variable is continuous, then it is a regression problem whereas when it contains discrete values, it is a classification problem.

##### **Logistic Regression:**

Logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring. An example of logistic regression could be applying machine learning to determine if a person is likely to be infected with COVID-19 or not.

## **Flask:**

Flask is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

## **Decision Tree:**

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

## **K- Nearest Classifier:**

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

## **Random Forest:**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.