# WEB PHISHING DETECTION

### PROJECT DEVELOPMENT PHASE

### SPRINT – 2

### PRE-REQUISITIES

**In order to develop this project we need to have the following requirements:**

## Project Hardware Requirements:-

- 2GB RAM (minimum)
- 100GB HDD (minimum)
- Intel 1.66 GHz Processor Pentium 4 (minimum)
- Internet Connectivity
- WINDOWS 7 or higher
- Python 3.6.0 or higher
- Visual Studio Code

# Project Software Requirements:

To build Machine learning models you must require the following packages

**Sklearn:** Scikit-learn is a library in Python that provides many unsupervised and supervised learning algorithms.

**NumPy:** NumPy is a Python package that stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object.

**Pandas:** pandas is a fast, powerful, flexible, and easy to use open source data analysis and manipulation tool,built on top of the Python programming language.

**Matplotlib:** It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits.

**Flask:** Web framework used for building Web applications.

**Logistic Regression:** Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

**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.

**HTML:** The HyperText Markup Language or HTML is the standard markup language for documents designed to be displayed in a web browser.