# **PROJECT: Build a detector for Phishing websites (KNN)**

The purpose of this document is to specify the requirements for the project “Build a detector for Phishing websites (KNN)” Apart from specifying the functional and non-functional requirements for the project, it also serves as an input for project scoping.

**Project Objective**

The purpose of the project is to use one or more of the classification algorithms to train a model on the Phishing website dataset in order to train a model which can detect phishing websites.

**Project Description and Scope**

You are provided with the following resources that can be used as inputs for your model:

1. A collection of website URLs for 11000+ websites. Each sample has 30 website parameters and then a class label identifying it as a phishing website or not (1 or -1).

2. Code template containing these code blocks :

a. Import modules (part 1)

b. Load data function + input/output field descriptions

You are expected to write the code for a binary classification model (phishing website or not) using Python Scikit-Learn that trains on the data and calculates the accuracy score on the test data.

**Project Guidelines**

Begin by extracting the ipynb file.

Exercise 1 : Build a phishing website classifier using KNN with no of neighbours as 5 and distance metric as “minkowski”.

Use 70% of data as training data and remaining 30% as test data.

(hint : use Scikit-Learn library KNeighboursClassifier)

(hint : look at the KNN tutorial taught earlier in the course)

Print count of misclassified samples in the test data prediction as well as accuracy score of the model.

**Pre-requisites**

To execute this project you will need:

* Python 3.5
* Scikit-Learn library
* Jupyter notebook

**Programming hint for the student :-**

For ideas, look at tutorials accompanying various modules taught in the course.