

## **TITLE : WEB PHISHING DETECTION**

### **Problem Statement:**

- Phishing detection techniques do suffer low detection accuracy and high false alarm especially when novel phishing approaches are introduced.
- Besides, the most common technique used, the blacklist-based method is inefficient in responding to emanating phishing attacks since registering a new domain has become easier, no comprehensive blacklist can ensure a perfect up-to-date database.
- Furthermore, page content inspection has been used by some strategies to overcome the false negative problems and complement the vulnerabilities of the stale lists.
- Moreover, page content inspection algorithms each have a different approach to phishing website detection with varying degrees of accuracy.

Therefore, an ensemble can be seen to be a better solution as it can combine the similarity in accuracy and different error-detection rate properties in selected algorithms. Therefore, this study will address a couple of research:

- How to process raw datasets for phishing detection?
- How to reduce the false negative rate in phishing websites algorithm?
- How to increase detection rate in phishing websites algorithms?
- What are the best compositions of classifiers that can give a good detection rate of phishing website

## **Abstract:**

- The chapter is organized as follows: first and foremost, a quick dive-in into the meaning of phishing in detail to enlighten the reader on why phishing is an important area of research is given;

- Second, different existing anti-phishing approaches are examined in terms of accuracy and limitations; third, a brief acknowledgment of existing techniques and how these techniques serve as a baseline to our research is presented.
- Furthermore, their advantages, as well as the setbacks experienced in the implementation of these techniques, are discussed. Fourth, we discuss the close technicalities of our work as implemented by other researchers in the same domain. This is also attributed to the basic knowledge behind the choice of algorithms and approaches used