Title : Web Phishing Detection

Problem Statement:

- Phishing detection technique do suffer low detection accuracy and high false alarm especially when novel phishing approaches are introduced.
- Besides, the most common technique used, blacklist-based method is inefficient in responding to emanating phishing attacks since registering 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 different approach to phishing website detection with varying degrees of accuracy. Therefore, 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 dataset for phishing detection?

- How to increase detection rate in phishing websites algorithms?
  .
- How to reduce false negative rate in phishing websites algorithm?

- What are the best compositions of classifiers that can give a good detection rate of phishing website

Abstact:

- The chapter is organized as follows: first and foremost, a quick dive-in to the meaning of phishing in details 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 also attributed to the basic knowledge behind the choice of algorithms and approaches used

Tools Used:

- A client-server framework was applied to exploit the superior detection performance of CBART and correspondingly conceal the computational overhead and memory requirement associated with it.
- The results demonstrate that the performance of potential classifiers at the clients, namely CART, SVM, NNet, and RF improves after using the predicted output of CBART in their datasets.
- CART achieves the maximum improvement in AUC of 1.49 percent. Despite the improvement in other classifiers—namely, RF by 0.17 percent, SVM by 0.06 percent, and NNet by 0.04 percent—the improvement in the AUC is apparently unnoticeable.