Problem Statement

Malicious links will lead to a website that often steals login credentials or financial information like credit card numbers. Attachments from phishing emails can contain malware that once opened can leave the door open to the attacker to perform malicious behavior from the user's computer.

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, blacklist-based method is inefficient in responding to emanating 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 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:

1.How to process raw dataset for phishing detection?

2.How to increase detection rate in Phishing Websites algorithms?

3.How to reduce false negative rate in phishing websites algorithm?

4.What are the best compositions of Classifiers that can give a good detection rate of Phishing sites?