A Methodical Overview on Phishing Detection along with an Organized Way to Construct an Anti-Phishing Framework

This paper focuses on the phishing detection on URL; which may be used together with other phishing detection as our dataset is the combination of phishing message and phishing URL.

It has several approaching, including:

1.**Heuristic Based Approach-** used heuristics like IP address in domain part, '@' symbol in URL,right click disabled, pop-up windows for passwords etc.
2.**Content Based Approach-** This technique makes use of Term Frequency/Inverse document Frequency (TF-IDF). TF-IDF compares the terms
in the original website to the phishy one.
3.**Blacklist Based Approach**- checking blacklists; the disadvantage is that it doesn't work out on new phishing URL
4.**Machine Learning Approach**- The one way focuses
5.**Hybrid Approach**- The combination of above approaches; considering the dataset is a combination of phishing URL & phishing message, we can further increase the accuracy of the classification with more methods

Process for machine learning:

Feature collection : It is not necessary if we have enough data from dataset

Feature selection: This process can be in the form of a feature matrix [3] [12] or feature list, MapReduce for feature extraction [5] or using dimensionality reduction; some features are listed in the tables 1 from paper

Classification: Various method can be used; such as Naive Bayes, SVM AdaBoost, J48, Random Forest etc

Conclusion: This paper discussed several ways to detect phishing, and it focuses on the URL; however, it doesn't have any special implementation on algorithm hence the feature selection becomes the most important task.