While classification is a supervised machine learning technique, clustering or cluster analysis is the opposite. It’s an unsupervised machine learning technique that you can use to detect similarities within an unlabeled dataset.

As per the process of spam filtering first terms get extracted by means of tokenization then through Local Concentration (LC) based extraction the frequency of words in datasets is calculated then this results feed as input to classifier. This method involve bag of words features to classify spam e-mail. this technique commonly used in text classification.