## Abstract

Clustering is one of data mining task that divides data into groups that elements in each group is similar between themself and are dissimilar to elements belonging to other groups. One of most used clustering algorithm is k-means.

In order to provide better clustering task, we need to process the data using Kernel Principal Component Analysis (KPCA). KPCA will produce a number of features (or called Principal Components) in greater numbers than the input data attribute. Then choose the number of PCs to be used based on the amount of variance. The greater variance owned by the PC, the PC is more represent the data. The next step is clustering using modified k-means. In modified k-means, the PC generated by KPCA will be used for deciding the initial centroid for k-means. Based on the test result, the number of features or PCs used in clustering process will affect the quality of cluster. In addition, the choice of kernel and kernel parameter values are used also affect the outcome of clustering.

Keywords : data mining, clustering, KPCA, modified k-means