Abstract

Clustering is one of data mining task. Clustering is grouping data by its characteristic into groups which commonly called clusters. The simplest algorithm for clustering is *k-means*. Eventhough k-means is considered a reliable and effective clustering algorithm, it still suffer several flaws. Those flaws are the input of number of clusters is still entered by user; and initialization of *centroids* still using random initialization.

These flaws on K-means are answered by Dan Pelleg and Andrew Moore, founders of X-means algorithm. X-means will recursively add the number of clusters until the existing number of clusters are same as the upper bound that entered by user. X-means final results are groups (clusters) of data; the best number of cluster based on the calculation when the program start.

This final task's intention is to implement X-means algorithm. Then, I calculate the accuracy of clusters using *silhouette coefficient* method. Finally, I analyze the result until there are conclusions that can be reached.

Keywords:
Clustering, K-means, *centroid*, X-means, *cluster range*, *Silhouette Coefficient*