Abstract

Data mining is a process to find an interesting pattern from very large data warehouse. In data mining, many tasks can be done. Begin from classification, clustering and rule association. In reality, the implementation of data mining is very much. For examples are churn prediction, fraud detection, customer segmentation etc. This paper discuss about customer segmentation using clustering method that called BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*).

"Clustering is the task of grouping the objects of a database into meaningful subclasses. So, it can answer the need of customer segmentation [8]". With segmentation, customers are divided into different segments, so the company can give a treat for their customers with different ways based on its segments. By customer segmentation, very large marketing database can be extracted in order to get important information about customer behavior and finally, the company can create an initiative or retention strategy.

BIRCH clustering method uses a highly specialized tree-structure for the purpose of clustering very large sets of point. Cluster BIRCH uses CF (*Cluster Feature*) concept which summaries the information about subclusters on CF Tree then clustered to k-groups using *clustering* hierarchical procedure. From the analysis, we can see that threshold as a parameter that inputed by user influences software's performance. The smaller threshold value, the number of subcluster created is larger and need more time to executes. Beside that, the increase the number of data and dimension influence preclustering and global clustering time, more data and dimension, so more time needed to execute the program because there are more subclusters created.

Keywords: data mining, customer segmentation, clustering, BIRCH