ABSTRACT

Clustering is one of data mining tech for grouping unlabeled dataset where the dataset has a similarity characteristics with the other in the same cluster and has a dissimilarity with other data in the another cluster. Data who has a big dissimilarity with other data called an outlier. This final task is focus to analisys Density-based Outlier Detection method, called Local Outlier Factor, to find an outlier in the dataset. One of outlier detection method is finding data density can be produce with Local Outlier Factor method that will count k-distance, k-distance neighborhood, reachability distance object, and local reachability density so that can produce LOF value that we can sort it. LOF will produce a degree of outlier for each object so that will decide whether the object is outlier or not. Local means it depend on how the object isolated from the surrounding neighborhood. If LOF value approaching 1, the data cannot be called local outlier. If the data has highest value of LOF, then we may call it an outlier. LOF of data depend on MinPts, that is a sum of nearest neighbors that used to define local neighborhood of an object.

Keywords : density-based outlier detection, local outlier factor