CLUSTER ENSEMBLE METHOD FOR CLUSTERING RURAL AREAS IN RIAU PROVINCE

Student Name : Angsoka Dewi
Register No. : 1310201713
Supervisor : Dr. Sutikno, S.Si., M.Si 
Co-supervisor : Dr. rer.pol. Heri Kuswanto, M.Si 

ABSTRACT

Rural areas present particular challenges on the implementation of census and surveys. Problem that often occurs in conducting these activities is sample replacement due to geographical constraints and remoteness. Therefore, the grouping of rural villages is needed in order to identify remote areas. The previous clustering algorithms focus on numerical or continuous data. However, datasets with mixed categorical and continuous data are common in real data. Clustering algorithm with mixed type variables is very limited and has not been available in any package. The purpose of this study is to assess and develop programs for cluster ensemble method dealing with mixed type variables. The program will be implemented for grouping rural villages in Riau Province. The clustering result using the cluster ensemble will be compared with clustering result using the full categorical and full continuous method. Comparison is done by looking at the minimum ratio of standard deviation within cluster and standard deviation between clusters. The results show that the ensemble method produce clusters with better performance than the full categorical as well as full continuous method. Ratios from each method measured under four clusters are 0.0072, 0.0904, and 0.2679 conservatively. By cluster ensemble, rural villages in Riau Province can be grouped into four clusters and it identifies 65 villages that are remote areas.

Keywords: Remote Area, Cluster Ensemble, Mixed Type Variables