

PENENTUAN JUMLAH *CLUSTER* OPTIMUM PADA *HIERARCHICAL CLUSTERING* MENGGUNAKAN *PSEUDO-F*

INTISARI

Kemiskinan merupakan suatu keadaan dimana seseorang tidak sanggup untuk memenuhi kebutuhan dirinya sendiri sesuai dengan taraf kehidupan lingkungannya. Berdasarkan data Badan Pusat Statistik, persentase penduduk miskin Sumatera Utara pada tahun 2021 mengalami kenaikan, yang sebelumnya 8,75% menjadi 9,01%. Nilai tersebut tentunya tidak menggambarkan keadaan kemiskinan tiap daerah. Artinya nilai tersebut hanya berlaku untuk Provinsi Sumatera Utara, sedangkan untuk tiap kabupaten/kota tidak berlaku demikian. Sumatera Utara juga merupakan provinsi dengan jumlah kabupaten/kota terbanyak ke-3 di Indonesia. Penelitian ini membahas bagaimana pemetaan wilayah Sumatera Utara berdasarkan faktor-faktor kemiskinan. Terdapat 10 variabel faktor-faktor kemiskinan yang digunakan sebagai dasar pengelompokan. Penelitian ini menggunakan analisis *hierarchical clustering*, dimana dilakukan perbandingan antara metode *single linkage*, *complete linkage*, dan *average linkage*. Pemilihan metode terbaik dilihat berdasarkan nilai *Pseudo-F statistics*. Dari 3 metode tersebut, didapatkan hasil pengelompokan paling optimum sebanyak 4 *cluster* dengan metode *complete linkage*. Setiap *cluster* memiliki karakteristik masing-masing. *Cluster* 1 yang beranggotakan Kota Medan dan Kota Pematang Siantar merupakan *cluster* yang memiliki tingkat kemiskinan yang sangat rendah. *Cluster* 2 merupakan *cluster* yang tingkat kemiskinannya termasuk dalam kategori rendah. *Cluster* 3 termasuk *cluster* dengan tingkat kemiskinan yang tinggi. Sedangkan *cluster* 4 merupakan *cluster* dengan tingkat kemiskinan sangat tinggi dengan anggota *cluster*-nya yaitu Kabupaten Nias Selatan dan Kabupaten Pakpak Bharat. Dengan demikian terlihat bahwa terjadi kesenjangan antar wilayah di Provinsi Sumatera Utara.

Kata Kunci: Provinsi Sumatera Utara, faktor kemiskinan, *hierarchical clustering*, *Pseudo-F statistics*

DETERMINING THE NUMBER OF OPTIMUM CLUSTERS IN HIERARCHICAL CLUSTERING USING PSEUDO-F

ABSTRACT

Poverty is a condition in which a person can't afford to fulfil his own needs appropriating to the living environment as standard. Based on data from Badan Pusat Statistik, the percentage of poor people in North Sumatra in 2021 has increased, from 8.75% to 9.01%. This value certainly doesn't describe the poverty condition of each region. Which means that this value only applies to North Sumatra Province, while for each regency/city this doesn't apply. North Sumatra is also the province with the 3rd largest number of regencies/cities in Indonesia. This study discusses how to map the North Sumatra region based on poverty factors. There are 10 variables of poverty factors that are used as the basis for grouping. This study uses hierarchical clustering analysis, where a comparison is made between single linkage, complete linkage, and average linkage methods. Selection of the best method is seen based on the value of Pseudo-F statistics. From these 3 methods, the most optimal grouping results were obtained in as many as 4 clusters with the complete linkage method. Each cluster has its characteristics. Cluster 1 which consists of Medan City and Pematang Siantar City is a cluster that has a very low poverty rate. Cluster 2 is a cluster with a low level of poverty. Cluster 3 is a cluster with a high poverty rate. While cluster 4 is a cluster with a very high poverty rate with cluster members namely South Nias Regency and Pakpak Bharat Regency. Thus it can be seen that there is a gap between regions in North Sumatra Province.

Keywords: North Sumatra Province, poverty factor, hierarchical clustering, Pseudo-F statistics