

RINGKASAN

PERKIRAAN JUMLAH KELOMPOK PADA ANALISIS KELOMPOK METODE *CENTROID* DAN *WARD* DENGAN MENGGUNAKAN *GAP STATISTIC*

Analisis kelompok (*Cluster analysis*) merupakan salah satu teknik analisis data multivariate yang bertujuan mengelompokkan beberapa obyek ke dalam kelompok sehingga obyek yang terletak pada satu kelompok memiliki sifat yang lebih dekat daripada obyek yang terletak di kelompok lain. Masalah utama yang sering muncul pada analisis kelompok adalah berapa jumlah kelompok (k) yang optimal dalam suatu himpunan data, terutama bila peneliti kurang memahami karakteristik himpunan data tersebut. Tibshirani, Wattner, dan Hastie (2001) mengusulkan metode *gap statistic* untuk menentukan jumlah kelompok optimum.

Indeks keluarga sejahtera merupakan salah satu hasil pengklasifikasian yang dapat diolah melalui teknik analisis kelompok. Berdasarkan penelitian Soenarnatalina (2006) keluarga sejahtera dapat dilihat dari beberapa indikator, yaitu : (1) Indikator kesehatan. (2) Indikator pendidikan. (3) Indikator perumahan dan lingkungan. (4) Indikator sosial budaya. (5) Indikator ekonomi.

Tujuan penelitian ini adalah : (1) Mengestimasi jumlah optimal kelompok pada metode pengelompokan *centroid* menggunakan *Gap Statistic* pada suatu himpunan data. (2) Mengestimasi jumlah optimal kelompok pada metode pengelompokan *ward* menggunakan *Gap Statistic* pada suatu himpunan data. (3) Membandingkan tingkat akurasi hasil estimasi jumlah kelompok dengan metode estimasi *Gap Statistic* (terhadap metode pengelompokan *centroid* dan *ward*).

Penelitian ini merupakan penelitian terapan dengan menggunakan data sekunder, yaitu hasil penelitian Soenarnatalina (2006) yang berjudul Pengembangan Indeks Keluarga Sejahtera di Provinsi Jawa Timur.

Hasil penelitian menunjukkan bahwa penggunaan *gap statistic* pada metode pengelompokan *centroid* diperoleh jumlah kelompok yang optimal adalah 5 kelompok. Dan penggunaan *gap statistic* pada metode pengelompokan *ward* diperoleh jumlah kelompok yang optimal adalah 5 kelompok. Sehingga dapat disimpulkan bahwa metode estimasi *gap statistic* dapat digunakan pada metode pengelompokan *centroid* dan *ward* karena memberikan hasil yang sama.

Berdasarkan hasil penelitian ini disarankan untuk membandingkan metode estimasi jumlah kelompok *gap statistic* dengan metode lainnya dengan sampel yang belum diketahui karakteristiknya.

ABSTRACT

ESTIMATING THE NUMBER OF CLUSTERS IN THE CLUSTER ANALYSIS OF CENTROID AND WARD METHODS USING THE GAP STATISTIC

One of the benefits of cluster analysis is to determine clusters from the available data. However, One of the most challenging problems in cluster analysis is choosing the optimal number of clusters in a dataset, mainly if the researcher less understands about the characteristic of group of data.

The purpose of this research was to: (1) estimate the number of optimum cluster in centroid clustering method using the gap statistic, (2) estimate the number of optimum cluster in ward clustering method using the gap statistic gap, (3) compare the accuracy level of group number estimation using the gap statistic method (to centroid and ward clustering methods).

This research was an applied research using secondary data from the result of the research done by Soenarnalita (2006) titled “Index development of wealthy family in East Java”. According to Soenarnalita (2006), there were five wealthy family clustering results based on the determined index, and those data were used as a standard to compare the using of group number estimation statistic gap method in both centroid and ward clustering procedures.

The result of this research showed that there were five optimum clusters when using statistic gap in centroid clustering method and five optimum cluster when using statistic gap in ward clustering method. It can be concluded that statistic gap estimation method can be used both in centroid and ward clustering data because its gives the same result.

Based on this result, it is suggested to compare the group number estimation using statistic gap method with other estimation methods for further research.

Keyword: Cluster analysis, Gap statistic