

INTISARI

Hidayah Budi Sayekti
19/445717/PA/19541

Analisis *cluster* adalah salah satu metode analisis multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Data yang dikumpulkan seiring waktu dapat disederhanakan menjadi bentuk yang lebih ringkas dengan mengubahnya menjadi fungsi atau kurva. Pengelompokan data fungsional menggunakan pendekatan probabilistik seperti *model based clustering* dilakukan dengan mengklasifikasikan data ke dalam *cluster* berdasarkan karakteristik fungsi atau kurva yang mewakili setiap data. Pembentukan fungsi dasar pada langkah awal dilakukan menggunakan pendekatan nonparametrik *cubic B-Spline*. Kemudian, algoritma *expectation maximization* digunakan dalam perhitungan probabilitas dan mengelompokkan data ke masing-masing *cluster*.

Metode ini diterapkan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Tengah tahun 2010-2022 berdasarkan indikator Indeks Pembangunan Manusia. Representasi data fungsional dilakukan dengan *cubic B-Spline basis smoothing* menggunakan 6 fungsi basis. Penentuan jumlah *cluster* dilakukan dengan menggunakan nilai *Bayesian Information Criterion* (BIC) terbesar. Hasil penelitian ini menunjukkan bahwa terdapat 4 *cluster* yang memiliki karakteristik berbeda. *Cluster* 1 terdiri dari 3 kota, *cluster* 2 terdiri dari 12 kabupaten/kota, *cluster* 3 terdiri dari 11 kabupaten/kota, dan *cluster* 4 terdiri dari 9 kabupaten/kota.

Kata kunci: analisis *cluster*, *model based clustering*, *cubic B-Spline*, algoritma EM

ABSTRACT

Hidayah Budi Sayekti
19/445717/PA/19541

Cluster analysis is one of the multivariate analysis methods that aims to group objects based on their characteristics. Data collected over time can be simplified into a more concise form by transforming it into a function or curve. Functional data clustering using probabilistic approaches such as model-based clustering is done by classifying data into clusters based on the characteristics of the function or curve that represents each data. Functional data representation in the first step is done using the cubic B-Spline nonparametric approach. Then, the expectation-maximization algorithm is used in calculating the probability and grouping the data into each cluster.

This method is applied to cluster regencies/cities in Central Java in 2010-2022 based on the Human Development Index indicator. Functional data representation is done with cubic B-Spline basis smoothing with 6 basis functions. The largest Bayesian Information Criterion (BIC) value is used to determine the number of clusters. In this study, four clusters are formed that have different characteristics. Cluster 1 consists of 3 cities, cluster 2 consists of 12 regencies/cities, cluster 3 consists of 11 regencies/cities, and cluster 4 consists of 9 regencies/cities.

Keywords: cluster analysis, model based clustering, cubic B-Spline, EM Algorithm