

INTISARI

OPTIMASI JUMLAH KLASTER PADA ALGORITMA K-MEANS MENGUNAKAN DAVIES BOULDIN INDEX

oleh

Nur Indah Setyaningsih

20/462323/PA/20295

Algoritma K-Means merupakan algoritma yang paling populer dan sering digunakan karena memiliki beberapa kelebihan seperti mudah dipelajari, efektif dan efisien, serta memiliki ketelitian yang cukup tinggi. Akan tetapi, algoritma K-Means tidak dapat menentukan jumlah klaster optimalnya sehingga diperlukan adanya metode yang dapat mengestimasi jumlah klaster optimal. Metode Davies Bouldin Index merupakan metode yang dapat digunakan untuk mengestimasi jumlah klaster optimal dengan menghitung rata-rata nilai rasio maksimum dari nilai kohesi yang dihitung dengan *sum of square within-cluster* dan separasi yang dihitung dengan *sum of square between-cluster*. Pada penelitian ini, data yang digunakan untuk studi kasus yaitu data banyaknya desa/kelurahan menurut keberadaan penyandang disabilitas di tiap kabupaten/kota Provinsi Jawa Timur tahun 2021. Berdasarkan analisis diperoleh bahwa analisis klaster menggunakan algoritma K-Means dengan jumlah klaster sebanyak 7 yang diestimasi menggunakan Davies Bouldin Index menghasilkan klaster yang paling optimal dengan nilai rasio *average distance between-cluster* dan *average distance within-cluster* sebesar 3,311868.

Kata kunci: Analisis Klaster, *K-Means*, Jumlah Klaster Optimal, Davies Bouldin Index

ABSTRACT

OPTIMIZATION NUMBER OF CLUSTER IN K-MEANS ALGORITHM USING DAVIES BOULDIN INDEX

by

Nur Indah Setyaningsih

20/4623232/PA/20295

The K-Means algorithm is a popular and frequently used clustering algorithm due to its ease of use, effectiveness, efficiency, and high accuracy. However, it cannot determine the optimal number of clusters, therefore a method that can estimate the optimal number of clusters is required. The Davies Bouldin Index method is used to estimate the optimal number of clusters by calculating the average maximum ratio value of the cohesion value, which is calculated by the sum of square within-cluster, and separation, which is calculated by the sum of square between-cluster. The data used for the case study is the number of villages/sub-districts according to the existence of persons with disabilities in each district/city of East Java Province in 2021. The analysis reveals that the K-Means algorithm, with a total of 7 clusters estimated using the Davies Bouldin Index, provides the most optimal cluster with a ratio value of the average distance between-cluster and the average distance within-cluster of 3.311868.

Keywords: Cluster Analysis, K-Means, Optimal Number of Cluster, Davies Bouldin Index