

INTISARI

KOMBINASI ANALISIS KLASSTER PARTISI DAN HIEARARKI DENGAN METODE BISECTING K-MEANS (STUDI KASUS: DATA TINGKAT PENGANGGURAN SELURUH KABUPATEN/KOTA DI PULAU JAWA TAHUN 2023)

Oleh

Rosidah Tamara

20/462328/PA/20300

Analisis klaster adalah metode multivariat yang dirancang untuk mengelompokkan objek berdasarkan karakteristik yang sama. Ada dua metode dalam analisis klaster, yaitu metode hierarki dan non-hierarki, atau metode partisi. Metode *bisecting k-means* mengkombinasikan metode partisi dengan metode hierarki dengan membagi satu klaster data menjadi dua pada setiap iterasinya. Skripsi ini membahas implementasi metode *bisecting k-means* untuk mengelompokkan data pengangguran seluruh kabupaten/kota di Pulau Jawa dengan variabel tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, indeks pembangunan masyarakat, dan persentase penduduk miskin. Hasil analisis klaster dengan metode *bisecting k-means* menghasilkan tiga klaster dengan performa hasil analisis yang lebih baik dibandingkan dengan metode *k-means*. Hal ini disimpulkan dari nilai empat metrik evaluasi yaitu *Sillhouette* sebesar 0.310466402, nilai *Davies-Bouldin Index* sebesar 1.085962392, dan nilai *WCSS* sebesar 259.2321922.

Kata kunci : analisis klaster, hierarki, bisecting k-means, k-means, tingkat pengangguran.

ABSTRACT

ENSAMBLE PARTITION AND HIERARCHICAL CLUSTERING WITH BISECTING K-MEANS METHODS (CASE STUDIED: UNEMPLOYMENT RATE DATA FOR ALL REGENCIES/CITIES IN JAVA ISLAND IN 2023)

By

Rosidah Tamara

20/462328/PA/20300

Cluster analysis is a multivariate method designed to group objects based on similar characteristics. There are two methods in cluster analysis, namely hierarchical and non-hierarchical methods, or partitioning methods. The bisecting k-means method combines the partitioning method with the hierarchical method by splitting one data cluster into two in each iteration. This thesis discusses the implementation of the bisecting k-means method to cluster unemployment data across all regencies/cities in Java Island using the variables of open unemployment rate, labor force participation rate, human development index, and the percentage of poor population. The cluster analysis results using the bisecting k-means method produced three clusters with better analysis performance compared to the k-means method. This conclusion is drawn from the values of four evaluation metrics: *Silhouette* score of 0.310466402, *Davies-Bouldin Index* score of 1.085962392, and *WCSS* score of 259.2321922.

Keywords : clustering, hierarchical, bisecting k-means, k-means, unemployment rate.