

ABSTRACT

CLUSTERING METHOD OF DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN) AND K-MEANS IN CLUSTERING PROVINCES BASED ON FACTORS THAT CAUSE STUNTING IN INDONESIA

By

NADIA ATHIYAH PUTRI

Clustering is a technique used to group data with similar characteristics within each cluster. There are several clustering methods available, such as K-Means and DBSCAN. K-Means is a clustering algorithm that partitions data into several clusters, while DBSCAN builds clusters based on connected densities. This research will use K-Means and DBSCAN clustering methods to perform clustering with stunting causal factor data and simulated data as a consideration of the performance of both methods. The results obtained on the data of the factors that cause stunting, the K-Means method has better performance based on the SI value of 0.5253 and the CH value of 26.0940. While in simulation data where the data is normally distributed and does not contain outliers, the DBSCAN method has better performance based on the SI value of 0.6495 and the CH value of 328.0302.

Kata kunci : Clustering, K-Means, DBSCAN, Stunting.

ABSTRAK

KLASTERISASI METODE *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* (DBSCAN) DAN *K-MEANS* DALAM PENGELOMPOKAN PROVINSI BERDASARKAN FAKTOR PENYEBAB *STUNTING* DI INDONESIA

Oleh

NADIA ATHIYAH PUTRI

Clustering adalah teknik yang digunakan untuk mengelompokkan data dengan karakteristik yang sama dalam setiap *cluster*. Ada beberapa metode *clustering* yang tersedia, seperti *K-Means* dan DBSCAN. *K-Means* adalah algoritma *clustering* yang mempartisi data ke dalam beberapa *cluster*, sedangkan DBSCAN membangun *cluster* berdasarkan densitas yang terhubung. Penelitian ini akan menggunakan metode *clustering K-Means* dan DBSCAN untuk melakukan *clustering* dengan data faktor penyebab *stunting* dan data simulasi sebagai pertimbangan performa dari kedua metode. Hasil yang diperoleh pada data faktor penyebab *stunting* metode *K-Means* memiliki performa yang lebih baik berdasarkan nilai SI sebesar 0.5253 dan nilai CH sebesar 26.0940. Sedangkan pada data simulasi dimana data tersebut berdistribusi normal dan tidak mengandung pencilan metode DBSCAN memiliki performa yang lebih baik berdasarkan nilai SI sebesar 0,6495 dan nilai *CH* sebesar 328,0302.

Kata kunci : *Clustering, K-Means, DBSCAN, Stunting.*