

ABSTRAK

Clustering dapat digunakan untuk mengetahui pola atau gambaran suatu informasi dalam jumlah cukup besar, seperti informasi kemiskinan di suatu wilayah. Di Indonesia, Badan Pusat Statistik (BPS) menginformasikan bahwa persentase kemiskinan di Indonesia per Maret 2023 mencapai 9,36%. Untuk mengetahui pola dan gambaran kemiskinan di Indonesia dapat menggunakan analisis *clustering* untuk mengelompokkan Kota/Kabupaten berdasarkan informasi kemiskinan di setiap wilayah (Fikri et al., 2021). Salah satu kendala dalam penerapan teknik *clustering* pada data ini adalah menentukan jumlah *cluster* yang paling sesuai untuk merepresentasikan perbedaan tingkat kemiskinan antar wilayah secara lebih akurat. Oleh karena itu, diperlukan proses pengelompokan data kemiskinan pada tingkat kabupaten atau kota dengan menguji berbagai nilai k , mulai dari 2 hingga 10.

Penelitian ini bertujuan untuk mengetahui kinerja yang lebih optimal antara algoritma *Bisecting K-Means* dan *K-Means* pada data informasi kemiskinan tahun 2023. Untuk menilai algoritma mana yang memiliki kinerja lebih baik dan optimal maka digunakan evaluasi *clustering* yaitu *Silhouette Coefficient* (SI). Hasil dari penelitian menunjukkan bahwa kinerja kedua algoritma yaitu setara karena memiliki nilai SI yang sama sebesar 0,7973 dengan jumlah $k = 2$. Namun berdasarkan kinerja *running time*, algoritma *Bisecting K-Means* lebih cepat daripada *K-Means*. Nilai *running time Bisecting K-Means* yaitu sebesar 0,0571 detik, sementara nilai *running time K-Means* sebesar 0,0687 detik. Penelitian ini juga dapat digunakan untuk memahami pola atau gambaran karakteristik kemiskinan dari suatu Kota/Kabupaten berdasarkan data informasi kemiskinan di Indonesia tahun 2023.

Kata kunci : *Clustering, Bisecting K-Means, K-Means, Silhouette Coefficient, running time*

ABSTRACT

Clustering can be used to determine the pattern or description of a large amount of information, such as poverty information in a region. In Indonesia, the Central Bureau of Statistics (BPS) informed that the percentage of poverty in Indonesia as of March 2023 reached 9,36%. To find out the pattern and description of poverty in Indonesia, we can use clustering analysis to group cities/districts based on poverty information in each region (Fikri et al., 2021). One of the obstacles in applying clustering techniques to this data is determining the most appropriate number of clusters to more accurately represent differences in poverty levels between regions. Therefore, it is necessary to cluster poverty data at the district or city level by testing various k values, ranging from 2 to 10.

This study aims to determine the more optimal performance between Bisecting K-Means and K-Means algorithms on poverty information data in 2023. To assess which algorithm has better and optimal performance, clustering evaluation is used, namely Silhouette Coefficient (SI). The results of the study show that the performance of the two algorithms is equivalent because they have the same SI value of 0,7973 with the number of $k = 2$. However, based on running time performance, the Bisecting K-Means algorithm is faster than K-Means. Bisecting K-Means running time value is 0,0571 seconds, while K-Means running time value is 0,0687 seconds. This research can also be used to understand the pattern or description of poverty characteristics of a city/district based on poverty information data in Indonesia in 2023.

Keywords : *Clustering, Bisecting K-Means, K-Means, Silhouette Coefficient, running time*