

INTISARI

PERBANDINGAN ANALISIS KLASTER MENGGUNAKAN *KMEANS* DAN *GENETIC ALGORITHM ON KMEANS (GA-KMEANS)*

Oleh

Arum Shelly Rahma

19/439208/PA/19031

Kmeans merupakan metode dalam analisis klaster yang mudah untuk diimplementasikan. Salah satu kelemahan dari *kmeans* adalah hasil klaster yang tidak konsisten dan selalu berubah-ubah. Hal ini disebabkan karena nilai *centroid* awal *kmeans* yang ditentukan secara acak sehingga menghasilkan *centroid* akhir yang berbeda-beda di setiap perulangan. Oleh karena itu, kualitas *kmeans* sangat bergantung terhadap nilai *centroid* awal. Salah satu metode pencarian nilai *centroid* yang optimal adalah algoritma genetika. Algoritma ini dikembangkan dari teori evolusi Darwin dan menerapkan prinsip-prinsip genetika dalam proses pencariannya, seperti seleksi, pindah silang, dan mutasi. Metode ini diterapkan untuk menggerombolkan produk minuman Starbucks UK pada musim semi (Februari-Mei 2023) berdasarkan kandungan nutriennya. Parameter yang digunakan dalam algoritma genetika adalah jumlah populasi sebanyak 100, jumlah generasi sebesar 100, teknik seleksi turnamen, teknik pindah silang dua titik dengan probabilitas 0,75, dan teknik mutasi *Gaussian* dengan probabilitas 0,001. Pada penelitian ini, *kmeans* dengan nilai *centroid* awal yang ditentukan oleh algoritma genetika atau *GA-Kmeans* lebih baik dan konsisten dibandingkan *kmeans* dengan *centroid* awal acak, dibuktikan dengan nilai SSE dan *Davies Bouldin Index* yang lebih rendah serta nilai *silhouette* dan *Calinski Harabaz* yang lebih tinggi. Namun, keterbatasan dari metode *GA-Kmeans* adalah waktu komputasi yang lebih tinggi dibandingkan *kmeans*.

Kata kunci: analisis klaster, *kmeans*, algoritma genetika, *centroid*

ABSTRACT

COMPARISON OF CLUSTER ANALYSIS USING KMEANS AND GENETIC ALGORITHM ON KMEANS (GA-KMEANS)

By

Arum Shelly Rahma

19/439208/PA/19031

Kmeans is a method in cluster analysis that is easy to implement. One of the drawbacks of kmeans is that the cluster results are not consistent and always changing. This is due to the initial centroid value kmeans is determined randomly so as to produce a final centroid different for each iteration. Therefore, the quality of the kmeans is highly dependent to the initial centroid value. One method of searching for centroid values optimal is the genetic algorithm. This algorithm is developed from the theory of evolution Darwin and applied genetic principles in his search process, such as of selection, crossover, and mutation. This method is applied to cluster Starbucks UK beverage products in the spring (February-May 2023) based on nutrient content. Parameters used in the genetic algorithm is the total population of 100, the number of generations of 100, the selection technique tournament, a two-point crossover technique with a probability of 0,75, and a mutation technique Gaussian with a probability of 0,001. In this study, kmeans with initial centroid values determined by the genetic algorithm or GA-Kmeans is better and consistent compared to kmeans with random initial centroids, evidenced by lower SSE and Davies Bouldin Index and higher Silhouette and Calinski Harabasz score. However, the limitation of the GA-Kmeans method is the higher computation time compared to kmeans.

Keywords: cluster analysis, kmeans, genetic algorithm, centroid