

Pengembangan algoritma K-means melalui pemilihan titik pusat awal = The development of K-means algorithm through the selection of initial centroid point

Aida Fastabiqah Khairati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20475274&lokasi=lokal>

Abstrak

ABSTRAK

Clustering merupakan proses pengelompokkan himpunan objek ke dalam cluster-cluster yang berisi objek-objek dengan kemiripan yang tinggi dibandingkan dengan objek-objek pada cluster lain. Algoritma K-means merupakan salah satu metode yang banyak digunakan dalam penyelesaian masalah clustering. Pada algoritma ini titik pusat cluster ditentukan oleh nilai rata-rata objek-objek pada cluster tersebut. Algoritma K-means memiliki ketergantungan terhadap pemilihan titik pusat awal cluster yang pemilihannya dilakukan secara acak. Pemilihan ini dapat mempengaruhi hasil clustering. Metode enhanced dan metode MMCA merupakan 2 contoh metode yang dapat diterapkan pada algoritma K-means dalam pemilihan titik pusat awal cluster. Penerapan kedua metode tersebut pada algoritma K-means menghasilkan hasil clustering yang lebih optimal, ditunjukkan dengan jumlah iterasi yang sama pada tiap simulasi untuk mencapai kriteria konvergen serta nilai rata-rata similaritas terhadap data benchmark yang lebih baik.

ABSTRACT

Clustering is the process of grouping the set of objects into clusters that contain objects with a high similarity compared to objects in other clusters. K means algorithm is one of many methods used in solving clustering problems. In this algorithm the center point of the cluster is determined by the average value of the objects in the cluster. The K means algorithm has a dependency on the selection of the cluster 39 s initial center point where the selection is randomly selected. This selection can affect the clustering result. The enhanced method and MMCA method are two examples of methods that can be applied to the K means algorithm in the selection of the cluster 39 s initial center point. The application of both methods to the K means algorithm resulted in a more optimal clustering result, indicated by the same number of iterations in each simulation to achieve convergent criteria as well as the average similarity value to better benchmark data.