

INTISARI

PENYELESAIAN MASALAH NILAI AWAL PUSAT KLASTER (CENTROID) PADA ANALISIS KLASTER K-MEANS DENGAN ALGORITMA PILLAR

Julius Satya Ratnandi

19/442595/PA/19344

K-Means adalah salah satu algoritma *clustering* yang populer digunakan karena algoritmanya mudah dipahami dan dapat digunakan untuk mengelompokkan data dengan cepat. Hasil pengelompokkan dengan K-Means dapat dipengaruhi dari pusat kluster (*centroid*) awal yang biasanya dipilih secara acak. Perbedaan titik *centroid* awal dapat memberikan hasil pengelompokkan yang berbeda. Sehingga, penting untuk memilih titik *centroid* awal yang dapat merepresentasikan anggota-anggota suatu kluster. Terdapat metode perhitungan untuk menentukan inisial *centroid* awal pada K-Means yakni algoritma *pillar*. Algoritma *pillar* terinspirasi dari peletakan pilar pada bagian sudut dan sisi dari gedung dan terletak berjauhan satu dengan lainnya. Pendekatan ini diimplementasikan pada kasus *clustering*, di mana titik *centroid* awal berada pada bagian luar dari kluster dan terletak berjauhan satu sama lain. Penelitian ini menggunakan 16 data sekunder yang memiliki karakteristik data yang berbeda-beda. Hasil inisialisasi *centroid* acak dan inisialisasi menggunakan algoritma *pillar* pada 16 data tersebut akan dibandingkan menggunakan indeks validasi *silhouette score* dan *davies-bouldin index* serta waktu eksekusi algoritma. Diperoleh kesimpulan hasil bahwa algoritma *pillar* dapat meningkatkan hasil dan mempercepat waktu eksekusi K-Means pada beberapa karakteristik data yang serupa dengan data yang digunakan.

Kata kunci : K-Means, analisis kluster, algoritma *pillar*, *centroid*, indeks validasi

ABSTRACT

SOLVING THE INITIAL CENTER VALUE (CENTROID) PROBLEM IN K-MEANS CLUSTERING USING PILLAR ALGORITHM

Julius Satya Ratnandi

19/442595/PA/19344

K-Means is one of the well known clustering algorithms used because the algorithm is easy to understand and can be used to group data quickly. The results of clustering with K-Means can be influenced from the initial cluster center (centroid) which is usually chosen randomly. Differences in the initial centroid points can give different grouping results. Thus, it is important to choose a starting centroid point that can represent the members of a cluster. There is a calculation method for determining the initial centroid in K-Means, namely pillar algorithm. The pillar algorithm is inspired by placing pillars at the corners and sides of the building and are located far from one another. This approach is implemented in clustering cases, where the initial centroid points are on the outside of the cluster and are far apart from each other. This study uses 16 secondary data which have different characteristics of statistical data values. The results of random centroid initialization and initialization using the pillar algorithm on the 16 data will be compared using the clustering validation index such as silhouette score and the davies-bouldin index as well as the execution time of the algorithm. It was concluded that the pillar algorithm can improve results and speed up K-Means execution time on several data characteristics that are similar to the data used.

Keywords : *K-Means, clustering, pillar algorithm, centroid, validation index*