

EFEKTIVITAS ALGORITMA *K-MEANS*, *K-MEDOIDS*, DAN OPTIMAL JUMLAH *CLUSTER* UNTUK MODIFIKASI ALGORITMA *K-MEDOIDS* PADA KUMPULAN DATA BESAR

Nur Aini Ikasakti

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

INTISARI

Kemajuan teknologi yang pesat mengakibatkan terus bertambahnya jumlah informasi. Kumpulan berbagai informasi dalam jumlah besar disebut sebagai *dataset*. Peningkatan jumlah informasi harus diiringi dengan penyimpanan yang baik. Namun, tanpa adanya pengorganisasian secara berkala dapat mengakibatkan penumpukan sehingga dokumen sulit untuk dicari, sebagai contoh ketika mencari dokumen yang memiliki tema yang sama akan membutuhkan waktu yang lama sehingga tidak efisien. Salah satu cara untuk menyelesaikan masalah tersebut menggunakan teknik *Clustering*. *K-Means* dan *K-Medoids* merupakan metode dalam teknik *clustering*. Metode *K-Means* dan *K-Medoids* sama-sama memiliki kelebihan yaitu, memiliki waktu komputasi yang cepat namun Metode *K-Means* memiliki kelemahan sensitif terhadap data *outlier*. Berbeda dengan Metode *K-Medoids* yang hampir dapat bekerja pada setiap jenis data matriks dan mampu mengatasi *outlier*. Penelitian ini akan membahas tentang efektivitas Modifikasi Algoritma *K-Medoids* untuk kumpulan data dalam jumlah besar dengan membandingkan rata-rata waktu komputasi dengan algoritma *K-Means*, dan *K-Medoids* yang diusulkan oleh Gandhi dan Sarivastava (2014). Dalam penelitian ini akan digunakan 1000 data yang dipilih secara *random* yang diambil dari data (*Internet Movie Database*) IMDB 5000 *Movie Dataset*. Hasil yang diperoleh yaitu Modifikasi Algoritma *K-Medoids* memiliki rata-rata waktu komputasi lebih cepat daripada Algoritma *K-Means* dan *K-Medoids*, sehingga Modifikasi Algoritma *K-Medoids* lebih efektif digunakan. Kemudian optimal jumlah *cluster* untuk modifikasi algoritma *K-Medoids* ada pada jumlah *cluster* 2, dengan menggunakan indeks Davies Bouldin (DB), indeks Calinski-Harabasz (CH), dan indeks Krzanowski-Lai (KL).

Kata Kunci : *Modifikasi Algoritma K-Medoids, Kumpulan Data Besar, Indeks Matrix*

EFFECTIVENESS OF K-MEANS, K-MEDOIDS, AND OPTIMAL THE NUMBER OF CLUSTERS TO MODIFIED K-MEDOIDS ALGORITHM FOR LARGE DATASET

Nur Aini Ikasakti

Statistics Department, Faculty of Mathematics and Natural Sciences

Universitas Islam Indonesia

ABSTRACT

Technological progress in resulting the increases of the information. A collection of various information in large numbers known as dataset. The increasing number of information must be accompanied by a good storage. But without any organization periodically can lead to difficulty in finding document, for instance when looking for document having the same theme will take a long time and inefficient. One way to deal with the problem is using clustering technique. K-Means and K-Medoids are engineering clustering methods. K-Means and K-Medoids method having same advantages namely, has time computing quickly but K-Means method have weakness that is sensitive to outlier data. In contrast, the K-Medoids almost work on any type of data matrix and to cope with outlier. This research going to talk about effectivity modification algorithms K-Medoids to a vast amounts of data by comparing average time computation of K-Means algorithm and K-Medoids proposed by Gopi Gandhi and Rohit Sarivastava. The research used 1000 data selected through random taken from the data of (the internet movie database) IMDB 5000 Movie Dataset. The results obtained namely K-Medoids modification algorithms having the average time computation faster than the algorithms K-Medoids and K-Means, so the K-Medoids modification algorithms is more effective to be used. Then the number of optimal clusters to modification algorithms K-Medoids is 2 cluster, with uses Davies Bouldin (DB) Index, Calinski-Harabasz Index (CH), and Krzanowski-Lai Index (KL) .

Keywords: *Modified K-Medoids Algorithm, Large Dataset, Matrix Index*