

ABSTRAK

Algoritma BIRCH dikenal karena kemampuannya menangani kumpulan data besar secara efisien. Hal ini ditunjukkan dalam sebuah studi yang membandingkan kinerja algoritma BIRCH dengan algoritma K-Means dan I-k-means. Dalam studi tersebut, hasilnya algoritma BIRCH mampu menghasilkan kualitas pengelompokan yang lebih baik dan running time yang lebih cepat dibandingkan dengan algoritma K-Means dan I-k-means dengan running time yang mirip dengan BIRCH, BIRCH cenderung menghasilkan kualitas pengelompokan yang lebih rendah. Besarnya jumlah fitur dataset menjadi salah satu faktor yang dapat memperburuk kualitas hasil pengelompokan. Menurut beberapa peneliti, algoritma pengelompokan sering kali kesulitan saat menangani data berdimensi tinggi. Untuk mengatasi masalah ini, studi ini membandingkan kinerja algoritma BIRCH dengan dan tanpa penerapan Principal Component Analysis (PCA), menggunakan kumpulan data Cover Type dari UCI Machine Learning Repository. Hasil pengelompokan dievaluasi menggunakan Adjusted Rand Index (ARI) dan metrik Purity untuk menentukan apakah PCA dapat meningkatkan kualitas pengelompokan algoritma BIRCH. Hasilnya menunjukkan bahwa penerapan PCA pada algoritma BIRCH tidak mampu meningkatkan kualitas pengelompokan.

ABSTRACT

The BIRCH algorithm is well known for its ability to efficiently handle large datasets. This was demonstrated in a study comparing the performance of BIRCH with K-Means and I-K-means. The results showed that BIRCH was able to produce better clustering quality and faster running time compared to K-Means. Although I-K-means achieved a runtime comparable to BIRCH, it tended to yield lower clustering quality. The high dimensionality of datasets is one factor that can worsen clustering performance. According to several researchers, clustering algorithms often struggle when dealing with high-dimensional data. To address this issue, the present study compares the performance of BIRCH with and without the application of Principal Component Analysis (PCA), using the Cover Type dataset from the UCI Machine Learning Repository. The clustering results were evaluated using the Adjusted Rand Index (ARI) and the Purity metric to determine whether PCA could enhance the clustering quality of BIRCH. The findings indicate that applying PCA to BIRCH did not improve its clustering quality.