

Kajian indeks validitas pada algoritma K-means = Study of validity indices on K-means algorithm

Amyra Aulia Adlina, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20474361&lokasi=lokal>

Abstrak

Indeks validitas merupakan metode yang mengevaluasi hasil clustering untuk mendapatkan jumlah kluster optimal suatu data. Pada skripsi ini, dilakukan clustering pada data menggunakan algoritma K-Means. Selanjutnya, hasil clustering tersebut dievaluasi oleh empat jenis indeks validitas, yaitu indeks Silhouette, indeks Davies-Bouldin, indeks Sum of Square, dan indeks Calinski-Harabasz. Implementasi keempat jenis indeks validitas dilakukan dengan menggunakan data benchmark yang sudah diketahui jumlah kelasnya. Hasil implementasi tersebut akan dibandingkan untuk mengetahui apakah keempat indeks validitas dapat memprediksi jumlah kluster dengan tepat. Dari hasil simulasi, indeks Silhouette, indeks Davies-Bouldin, dan indeks Calinski-Harabasz dapat memprediksi jumlah kluster optimal lebih tepat dibandingkan dengan indeks Sum of Square.

.....The validity index is a method that evaluates the clustering results to get the optimal number of clusters of a data. In this skripsi, data clustered using K Means algorithm. Furthermore, the clustering results are evaluated by four types of validity indices, namely the Silhouette index, the Davies Bouldin index, the Sum of Square index, and the Calinski Harabasz index. The implementation of the four validity indices is done by using the benchmark data which is already known to the number of classes.

The results of the implementation will be compared to find out whether the four validity indices can predict the number of clusters appropriately. From the simulation results, the Silhouette index, the Davies Bouldin index, and the Calinski Harabasz index can predict the optimal cluster number is more precise than the Sum of Square index.