

INTISARI

Clustering merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. Penelitian ini membandingkan metode K-Means dan Fuzzy C-Means dalam pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan data yang diperoleh dari Badan Pusat Statistik (BPS) Jawa Tengah tahun 2023. Analisis dilakukan melalui beberapa tahapan, termasuk normalisasi data, seleksi fitur, serta penentuan jumlah cluster optimal menggunakan Metode Elbow, yang menghasilkan jumlah cluster optimal sebanyak 3. Evaluasi hasil clustering dilakukan menggunakan Silhouette Coefficient dan Davies-Bouldin Index (DBI) untuk menilai kualitas hasil pengelompokan. Hasil penelitian menunjukkan bahwa K-Means menghasilkan Silhouette Coefficient sebesar 0,620 dan DBI sebesar 0,461, sedangkan Fuzzy C-Means menghasilkan Silhouette Coefficient sebesar 0,619 dan DBI sebesar 0,461. Meskipun hasil keduanya hampir serupa, K-Means menunjukkan sedikit keunggulan dalam konsistensi hasil pengelompokan dan efisiensi komputasi. Selain itu, proses seleksi fitur dan normalisasi data juga tidak berpengaruh meningkatkan kualitas clustering yang dihasilkan. Kesimpulan dari penelitian ini menunjukkan bahwa baik K-Means maupun Fuzzy C-Means dapat digunakan untuk clustering wilayah berdasarkan karakteristik yang dianalisis, dengan K-Means lebih direkomendasikan untuk penerapan yang membutuhkan efisiensi tinggi. Hasil penelitian ini diharapkan dapat menjadi referensi bagi analisis clustering berbasis data di berbagai bidang.

Kata kunci: Clustering, K-Means, Fuzzy C-Means, Data Mining, Jawa Tengah

ABSTRACT

Clustering is one of the techniques in data mining used to group data based on similarities in characteristics. This research compares the K-Means and Fuzzy C-Means methods in clustering districts/cities in Central Java Province based on data obtained from the Central Statistics Agency (BPS) of Central Java in 2023. The analysis is conducted through several stages, including data normalization, feature selection, and determining the optimal number of clusters using the Elbow Method, which results in an optimal number of clusters of 3. The evaluation of the clustering results is performed using the Silhouette Coefficient and Davies-Bouldin Index (DBI) to assess the quality of the clustering results. The research findings indicate that K-Means produces a Silhouette Coefficient of 0.620 and a DBI of 0.461, while Fuzzy C-Means produces a Silhouette Coefficient of 0.619 and a DBI of 0.461. Although the results of both are nearly similar, K-Means demonstrates a slight advantage in the consistency of clustering results and computational efficiency. Furthermore, the feature selection process and data normalization do not significantly enhance the quality of the resulting clustering. The conclusion of this study indicates that both K-Means and Fuzzy C-Means can be utilized for clustering regions based on the analyzed characteristics, with K-Means being more recommended for applications that require high efficiency. The findings of this research are expected to serve as a reference for data-based clustering analysis in various fields.

Keywords: Clustering, K-Means, Fuzzy C-Means, Data Mining, Central Java