

ABSTRAK

Produksi padi di Indonesia memiliki karakteristik yang berbeda antarwilayah dan antarperiode pengamatan. Perbedaan tersebut dapat dilihat berdasarkan curah hujan, luas panen, dan produksi padi sehingga diperlukan analisis untuk mengidentifikasi kelompok data yang memiliki karakteristik serupa. Penelitian ini bertujuan untuk membandingkan algoritma *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* dan *Agglomerative Hierarchical Clustering (AHC)* dalam mengelompokkan data produksi padi di Indonesia serta menganalisis karakteristik *cluster* yang terbentuk. Data yang digunakan merupakan data bulanan tahun 2025 dari 38 provinsi dengan 456 data yang diperoleh dari Badan Pusat Statistik (BPS) dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Variabel yang digunakan meliputi curah hujan, luas panen, dan produksi padi, sedangkan variabel bulan ditransformasikan menggunakan *cyclical encoding* menjadi *Bulan_sin* dan *Bulan_cos*. Data kemudian distandarisasi menggunakan *StandardScaler*. Parameter DBSCAN ditentukan melalui *K-Distance Graph* dan *Grid Search*, sedangkan AHC diuji menggunakan beberapa metode *linkage*. Hasil penelitian menunjukkan bahwa AHC dengan *Complete Linkage* menghasilkan kualitas *clustering* yang lebih baik, dengan 2 *cluster*, nilai *Silhouette Coefficient* sebesar 0,7046, dan *Davies-Bouldin Index* sebesar 0,3464. Sementara itu, DBSCAN menghasilkan 2 *cluster* dengan 17 data *noise*, nilai *Silhouette Coefficient* sebesar 0,5008, dan *Davies-Bouldin Index* sebesar 0,6221. Analisis karakteristik menunjukkan bahwa pada kedua algoritma terdapat kecenderungan bahwa kelompok dengan luas panen lebih besar memiliki karakteristik produksi yang lebih tinggi. Berdasarkan hasil evaluasi, AHC dengan *Complete Linkage* merupakan metode yang memberikan hasil *clustering* terbaik pada penelitian ini.

Kata kunci: Produksi Padi, *Clustering*, DBSCAN, *Agglomerative Hierarchical Clustering*, *Silhouette Coefficient*, *Davies-Bouldin Index*.

ABSTRACT

Rice production in Indonesia has different characteristics across regions and observation periods. These differences can be observed based on rainfall, harvested area, and rice production, requiring an analysis to identify groups of data with similar characteristics. This study aims to compare the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Agglomerative Hierarchical Clustering (AHC) algorithms for clustering rice production data in Indonesia and analyzing the characteristics of the resulting clusters. The data used consist of monthly data for 2025 from 38 provinces with 456 data, obtained from the Central Statistics Agency (BPS) and the Meteorology, Climatology, and Geophysical Agency (BMKG). The variables used include rainfall, harvested area, and rice production, while the month variable was transformed using cyclical encoding into Month_sin and Month_cos. The data were then standardized using StandardScaler. DBSCAN parameters were determined using the K-Distance Graph and Grid Search, while AHC was evaluated using several linkage methods. The results show that AHC with Complete Linkage produced better clustering quality, resulting in 2 clusters with a Silhouette Coefficient of 0.7046 and a Davies-Bouldin Index of 0.3464. Meanwhile, DBSCAN produced 2 clusters with 17 noise observations, a Silhouette Coefficient of 0.5008, and a Davies-Bouldin Index of 0.6221. The cluster characteristic analysis indicates that both algorithms show a tendency for groups with a larger harvested area to have higher production characteristics. Based on the evaluation results, AHC with Complete Linkage is the method that provides the best clustering results in this study.

Keywords: Rice Production, Clustering, DBSCAN, Agglomerative Hierarchical Clustering, Silhouette Coefficient, Davies-Bouldin Index.