

ME3202 – Analisis Data Cuaca dan Iklim II

Praktikum Modul 6: *Cluster Analysis*



Dosen:

Muhammad Ridho Syahputra, S.Si., M.Si.

Disusun Oleh:

12821046 – Fardhan Indrayesa

Kode Kelas: ME3202

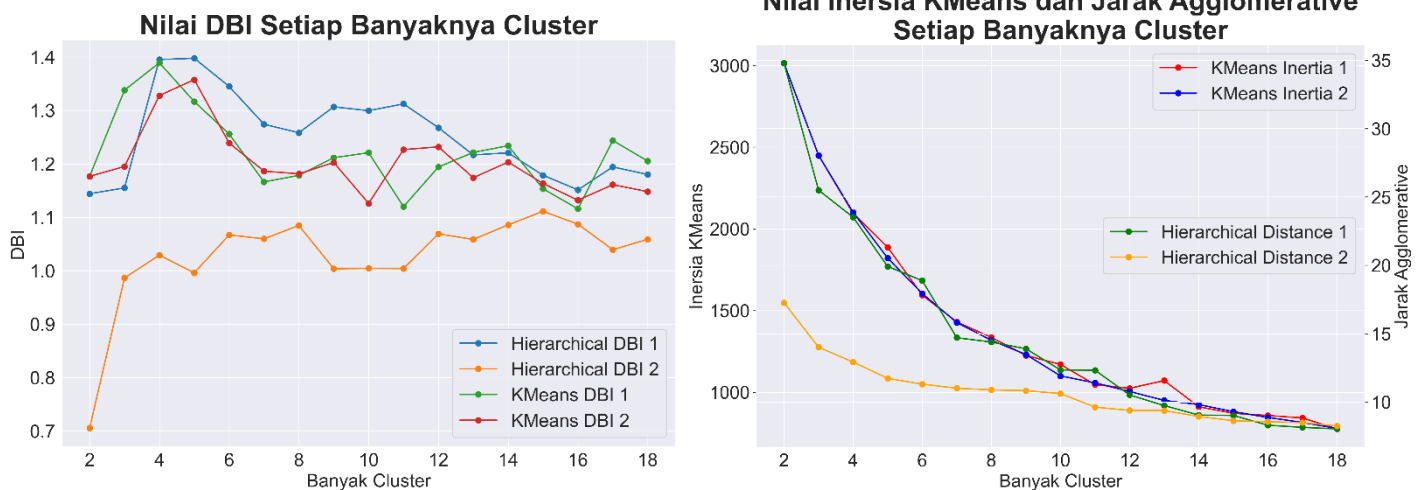
**Program Studi Meteorologi
Fakultas Ilmu dan Teknologi Kebumihan
Institut Teknologi Bandung
2021**

Data yang digunakan adalah data *precipitation rate* ([link data](#)) per bulan dengan satuan mm/s di seluruh dunia dari tanggal 1 Januari 1979 – 1 April 2024. Data dipotong menjadi area Indonesia dari tanggal 1 Januari 1991 – 1 Desember 2020 (30 tahun) dan dikonversi menjadi satuan mm/hari. Variasi pasangan metode dan *linkage* yang digunakan untuk *hierarchical clustering* adalah euclidean ward dan manhattan average. Evaluasi yang digunakan adalah metode elbow (inersia Kmeans dan jarak Agglomerative) dan Davies-Bouldin Index (DBI). Banyak cluster yang dipilih adalah 5 karena memiliki nilai indeks DBI yang tidak terlalu tinggi pada beberapa variasi.

Variasi yang digunakan:

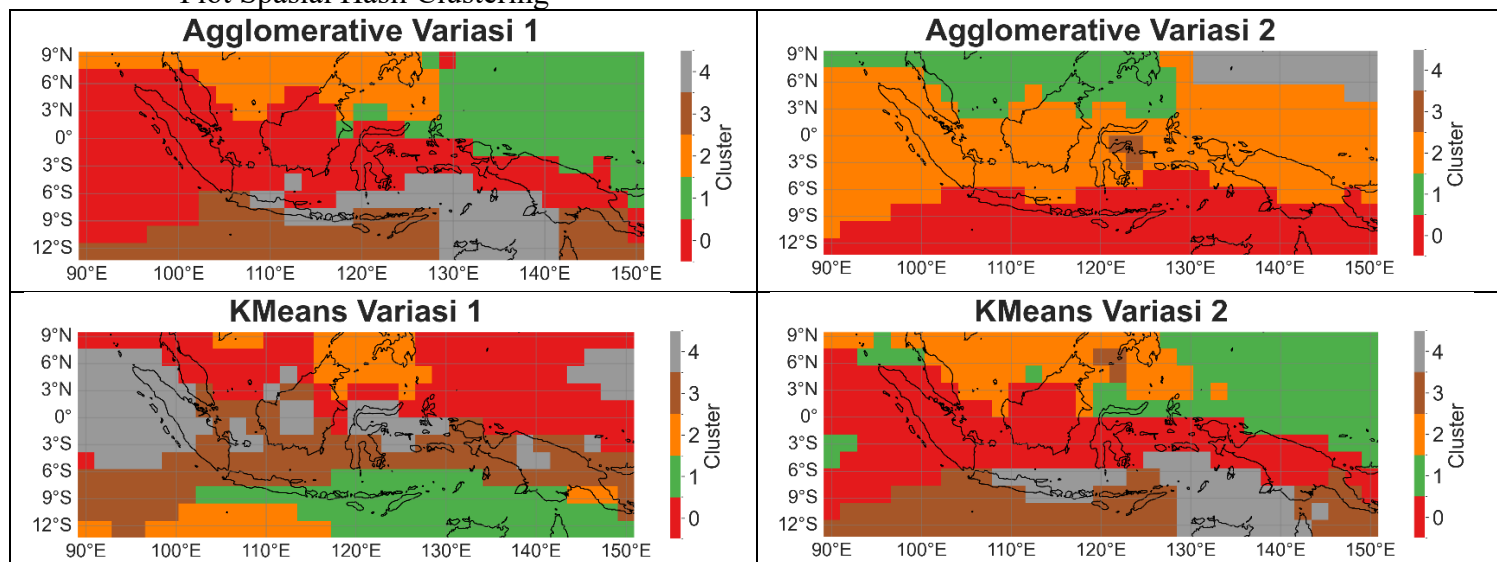
- Agglomerative variasi 1: Agglomerative, metode euclidean, linkage ward
- Agglomerative variasi 2: Agglomerative, metode manhattan, linkage average
- KMeans variasi 1: inisialisasi kmeans++
- KMeans variasi 2: inisialisasi data sentroid Agglomerative variasi 1.

- Evaluasi Hasil Clustering



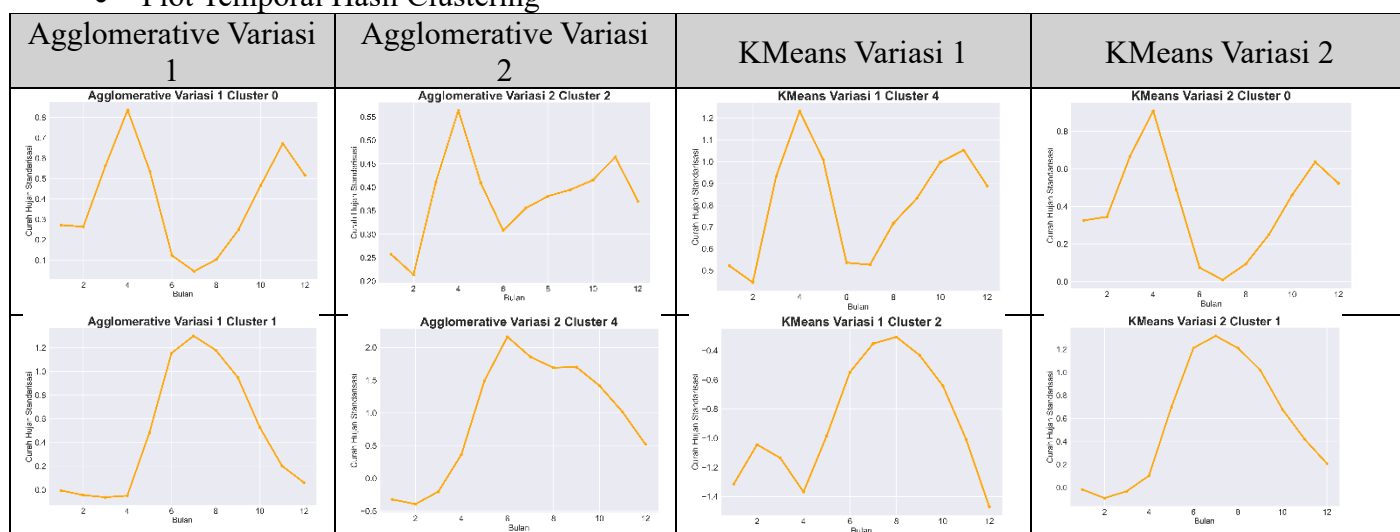
Gambar di atas adalah grafik nilai DBI, inersia KMeans, dan jarak Agglomerative setiap banyaknya cluster yang digunakan. Nilai DBI memiliki nilai *perfect* 0. Berdasarkan gambar kiri (DBI), terlihat bahwa variasi yang memiliki nilai DBI terendah adalah variasi Agglomerative 2 dengan banyak cluster 2. Untuk banyak cluster 5, nilai DBI yang paling rendah dihasilkan oleh variasi Agglomerative 2, dan nilai DBI yang paling tinggi dihasilkan oleh variasi Agglomerative 1. Hal ini menandakan bahwa hasil cluster dengan variasi Agglomerative 2 memiliki jarak antar anggota cluster lebih dekat dan jarak antar cluster yang lebih jauh dibandingkan dengan variasi yang lain. Selain itu, gambar kanan adalah plot inersia KMeans dan jarak Agglomerative setiap banyaknya cluster. Semakin besar nilai inersia KMeans, maka jarak antar anggota cluster semakin renggang, semakin besar nilai jarak Agglomerative, perbedaan antar cluster semakin jelas. Jarak 5 cluster untuk Agglomerative variasi 2 memiliki nilai yang paling kecil dan jarak Agglomerative variasi 1 memiliki nilai yang paling besar, yang menandakan bahwa Agglomerative variasi 1 memiliki perbedaan yang tinggi antar clusternya. Inersia 5 cluster untuk KMeans variasi 1 memiliki nilai yang lebih tinggi daripada inersia KMeans variasi 2, yang menandakan bahwa KMeans variasi 2 memiliki jarak antar anggota cluster yang lebih rapat daripada KMeans variasi 1.

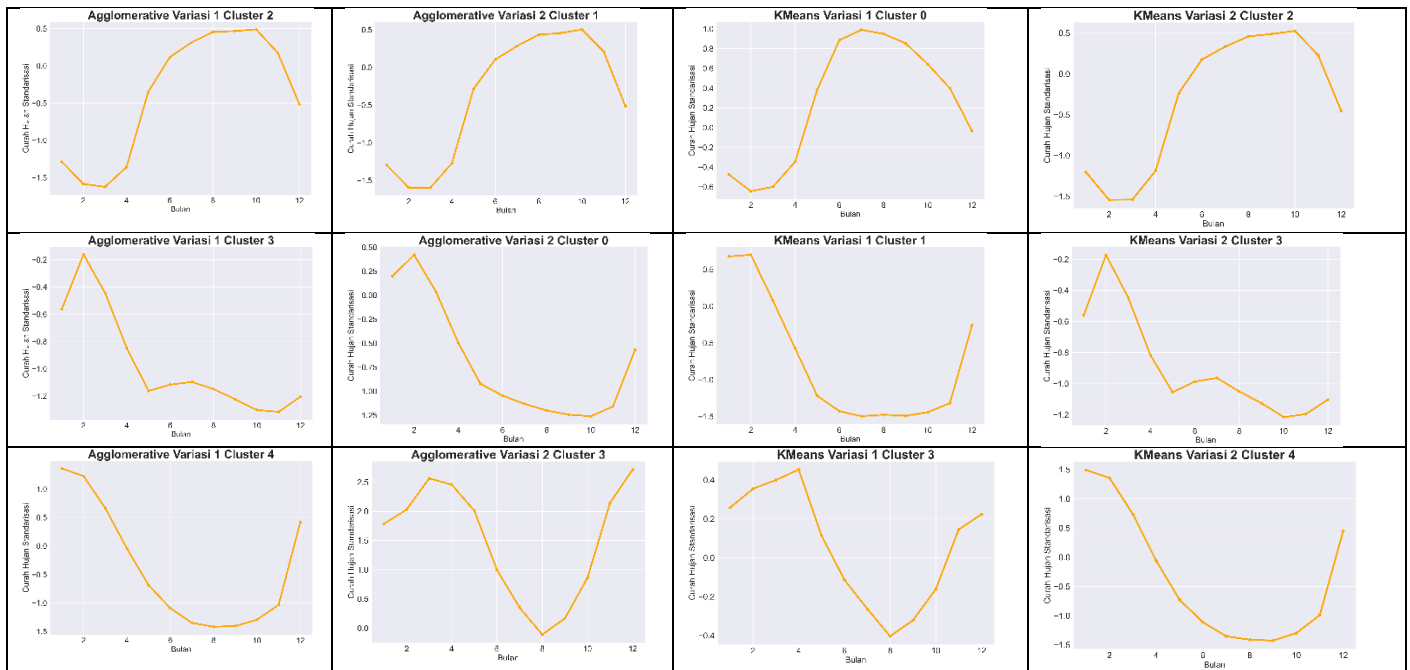
- Plot Spasial Hasil Clustering



Gambar di atas adalah plot spasial label setiap variasi clustering. Berdasarkan gambar di atas, terlihat bahwa hasil clustering Agglomerative variasi 1 dan KMeans variasi 2 memiliki persebaran cluster yang hampir sama karena pada KMeans variasi 2, data inisialnya menggunakan hasil clustering Agglomerative variasi 1. Untuk Agglomerative variasi 1, cluster yang paling dominan adalah cluster 0 dan 4. Cluster yang dominan untuk Agglomerative variasi 2 adalah cluster 0 dan 2. Cluster yang dominan untuk KMeans variasi 1 adalah cluster 1, 3, dan 4. Cluster yang dominan untuk KMeans variasi 2 adalah cluster 0 dan 4. Hasil cluster dapat berbeda-beda karena metode clustering dan metode keserupaan yang digunakan juga berbeda-beda.

- Plot Temporal Hasil Clustering





Gambar di atas adalah grafik deret waktu pola curah hujan per bulan untuk setiap variasi clustering dan setiap cluster. Pola curah hujan pada Agglomerative variasi 1 dan KMeans memiliki pola yang cukup mirip karena data inisial KMeans variasi 2 menggunakan data hasil cluster Agglomerative variasi 1. Berdasarkan gambar di atas untuk seluruh metode clustering, terlihat bahwa:

- Cluster 0 pada Agglomerative 1 dan KMeans 2, cluster 2 pada Agglomerative 2, dan cluster 4 pada KMeans 1 memiliki pola yang serupa dan memiliki dua puncak curah hujan, yaitu di bulan April dan bulan November.
- Cluster 1 pada Agglomerative 1 dan KMeans 2, cluster 4 pada Agglomerative 2, dan cluster 2 pada KMeans 1 memiliki pola yang serupa dan memiliki satu puncak curah hujan, yaitu di sekitar bulan Juni hingga Agustus.
- Cluster 2 pada Agglomerative 1 dan KMeans 2, cluster 1 pada Agglomerative 2, dan cluster 0 pada KMeans 1 memiliki pola yang serupa dan memiliki satu puncak curah hujan, yaitu di sekitar bulan Juli hingga Oktober.
- Cluster 3 pada Agglomerative 1 dan KMeans 2, cluster 0 pada Agglomerative 2, dan cluster 1 pada KMeans 1 memiliki pola yang serupa dan memiliki satu puncak curah hujan, yaitu di bulan Februari.
- Cluster 4 pada Agglomerative 1 dan KMeans 2, cluster 3 pada Agglomerative 2 dan KMeans 1 memiliki pola yang serupa dan memiliki dua puncak curah hujan, yaitu di sekitar bulan Januari hingga April dan di bulan Desember.