

Indonesia AI

Clustering Algorithm

Proprietary document of Indonesia AI 2023



OBJECTIVE & OUTLINE

Proprietary document of Indonesia AI 2023



Clustering Algorithm

Objektif: Memahami konsep dasar algoritma clustering yang merupakan bagian dari metode unsupervised learning.

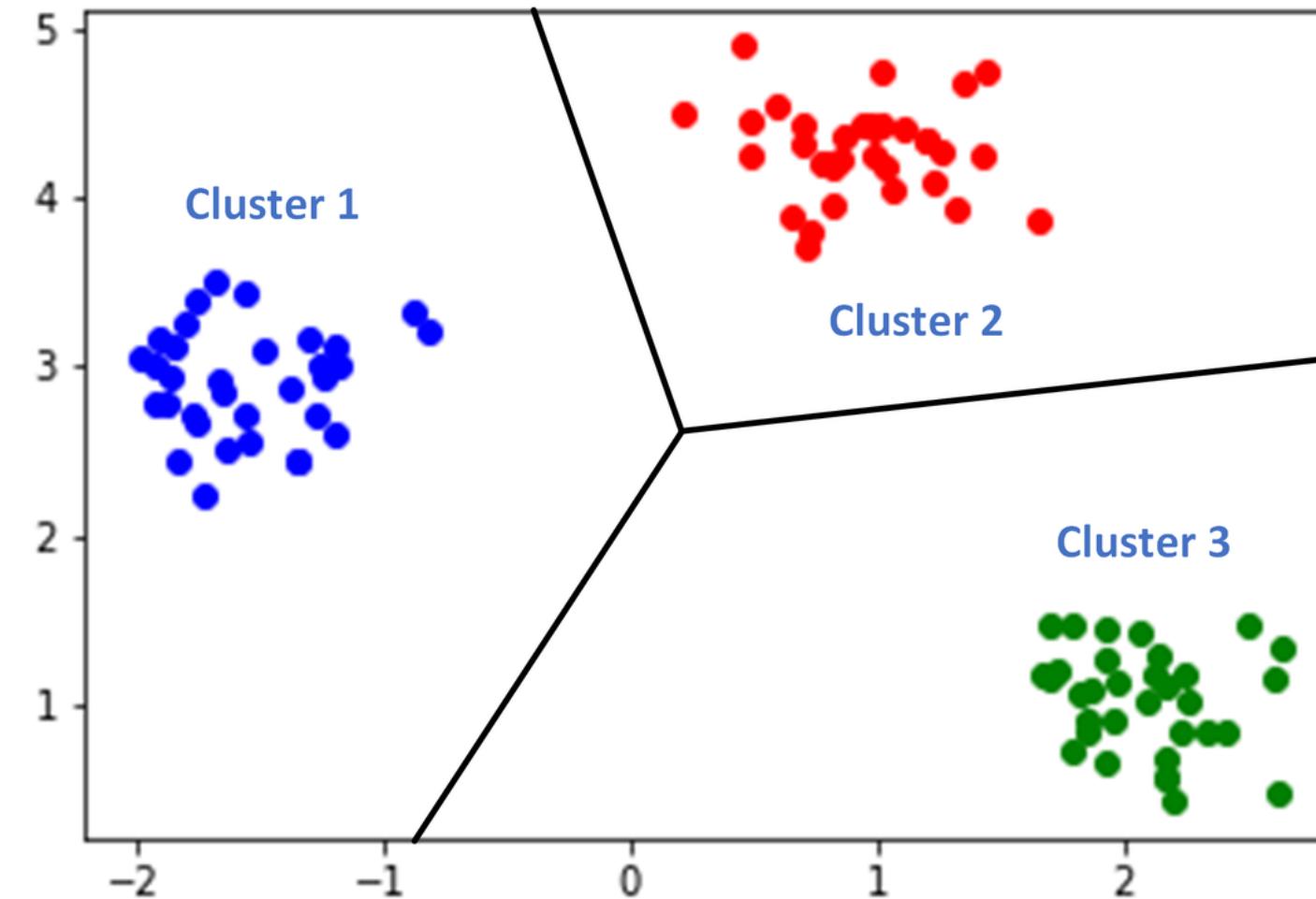
Outline:

1. Clustering Concept
2. Algoritma K-Means Clustering
3. Algoritma Hierarchical Clustering
4. Algoritma DBSCAN

— Clustering Concept

DEFINISI

Proprietary document of Indonesia AI 2023



Clustering adalah teknik dalam algoritma Machine Learning yang digunakan untuk mengelompokkan data menjadi beberapa kelompok atau cluster, berdasarkan kesamaan karakteristik (similarity) di antara titik data.

TUJUAN

Proprietary document of Indonesia AI 2023

Tujuan utama dari clustering adalah untuk membagi data ke dalam **kelompok cluster** yang berbeda sehingga data dalam satu kelompok memiliki kesamaan (similarity) yang lebih besar dengan satu sama lain daripada dengan data di **kelompok cluster** lain.

Indonesia AI

Image source: www.codeproject.com

TASK CLUSTERING DI COMPUTER VISION

Proprietary document of Indonesia AI 2023

Beberapa Task Clustering di Computer Vision adalah:

1. Color Clustering (Adaptive Threshold)
2. Vector Embedding Clustering
3. Text Clustering on OCR



Indonesia AI

Image source: www.codeproject.com

Algoritma K-Means Clustering

DEFINISI

Proprietary document of Indonesia AI 2023



Algoritma K-Means Clustering adalah salah satu algoritma clustering yang paling populer dalam Aplikasi Machine Learning.

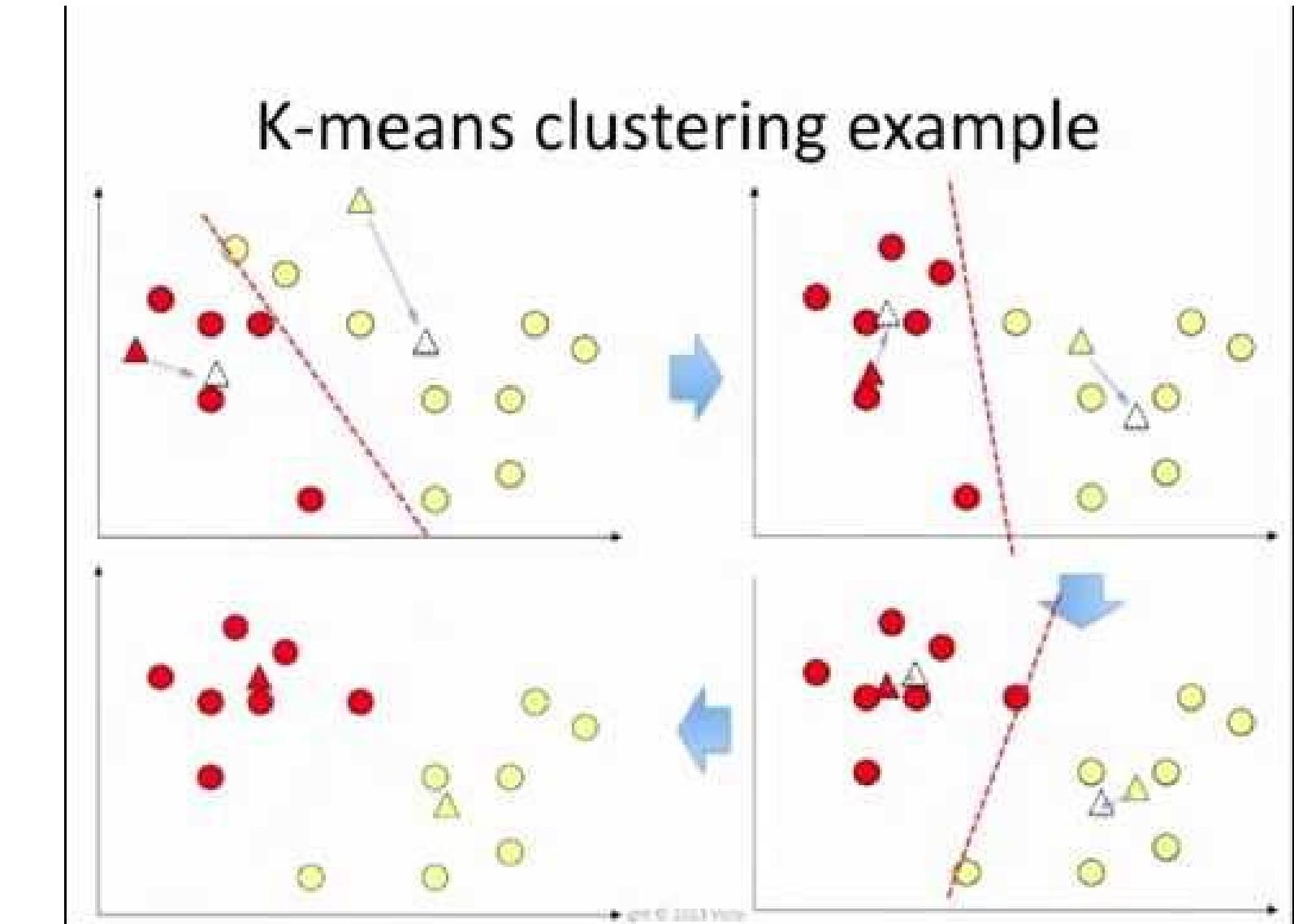
Algoritma K-Means adalah metode pengelompokan data unsupervised learning yang digunakan untuk membagi data menjadi kelompok yang berbeda atau cluster.

DEFINISI

Proprietary document of Indonesia AI 2023

Pada dasarnya, algoritma K-Means mencoba untuk menemukan pusat dari masing-masing kelompok atau cluster dan mengelompokkan data ke dalam kelompok yang memiliki pusat yang paling dekat dengan setiap data.

Proses ini berulang sampai pusat dari setiap kelompok stabil, dan data tidak berpindah dari satu kelompok ke kelompok lainnya.



Secara singkat, algoritma K-Means bekerja dengan cara sebagai berikut:

1. Menentukan **jumlah kelompok atau cluster** yang diinginkan



Secara singkat, algoritma K-Means bekerja dengan cara sebagai berikut:

2. Menentukan **titik awal pusat (centroid)** untuk setiap kelompok secara acak



Secara singkat, algoritma K-Means bekerja dengan cara sebagai berikut:

3. Menghitung **jarak antara setiap titik data** dengan setiap pusat kelompok



Secara singkat, algoritma K-Means bekerja dengan cara sebagai berikut:

4. **Menetapkan setiap titik data** ke kelompok atau cluster yang memiliki pusat (centroid) terdekat



Secara singkat, algoritma K-Means bekerja dengan cara sebagai berikut:

5. **Menghitung ulang pusat (centroid)** dari setiap kelompok atau cluster berdasarkan titik-titik data yang telah ditetapkan ke dalam kelompok atau cluster tersebut



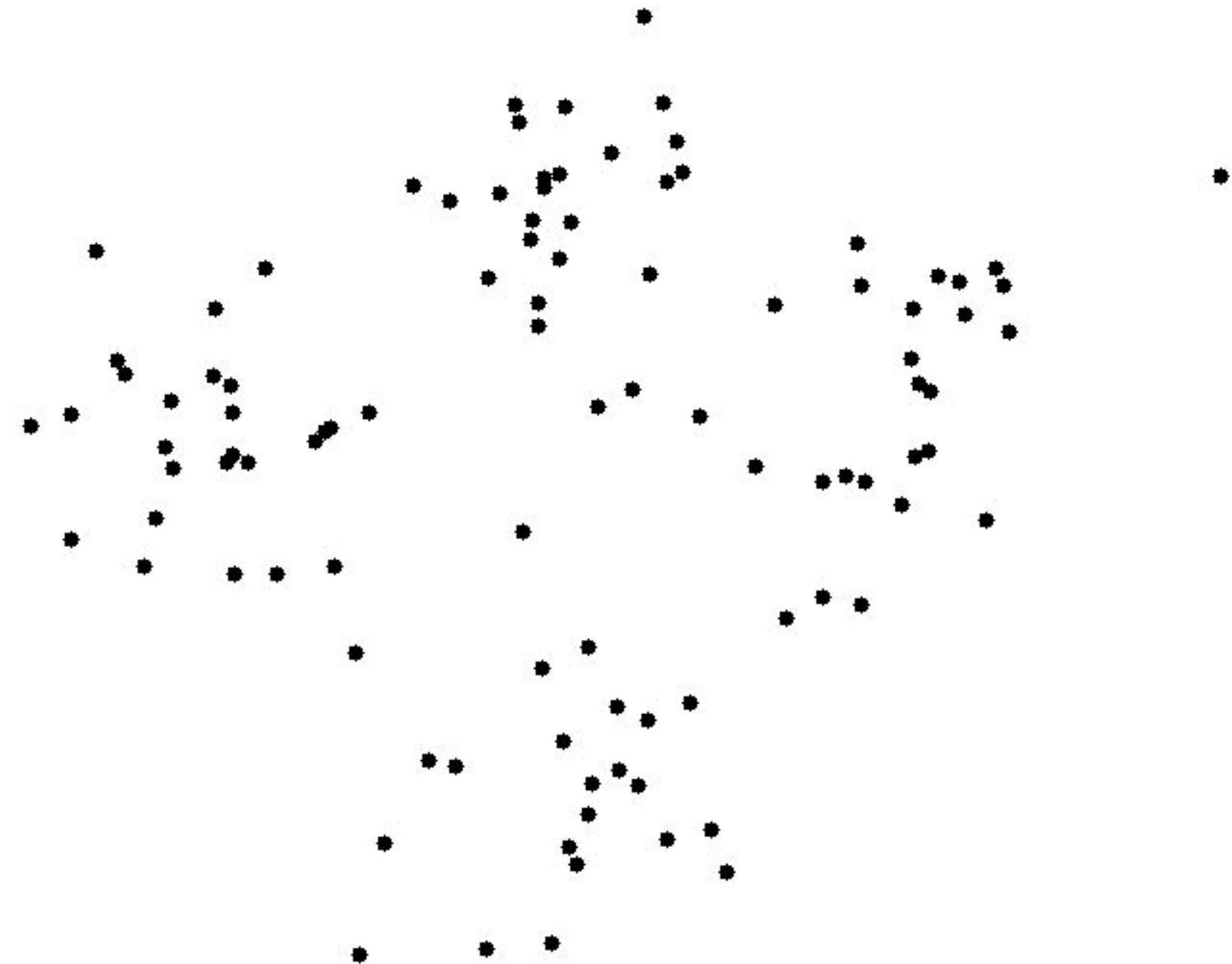
Secara singkat, algoritma K-Means bekerja dengan cara sebagai berikut:

6. **Mengulang langkah 3 sampai 5** sampai titik pusat (centroid) stabil atau tidak berubah



ILUSTRASI

Proprietary document of Indonesia AI 2023



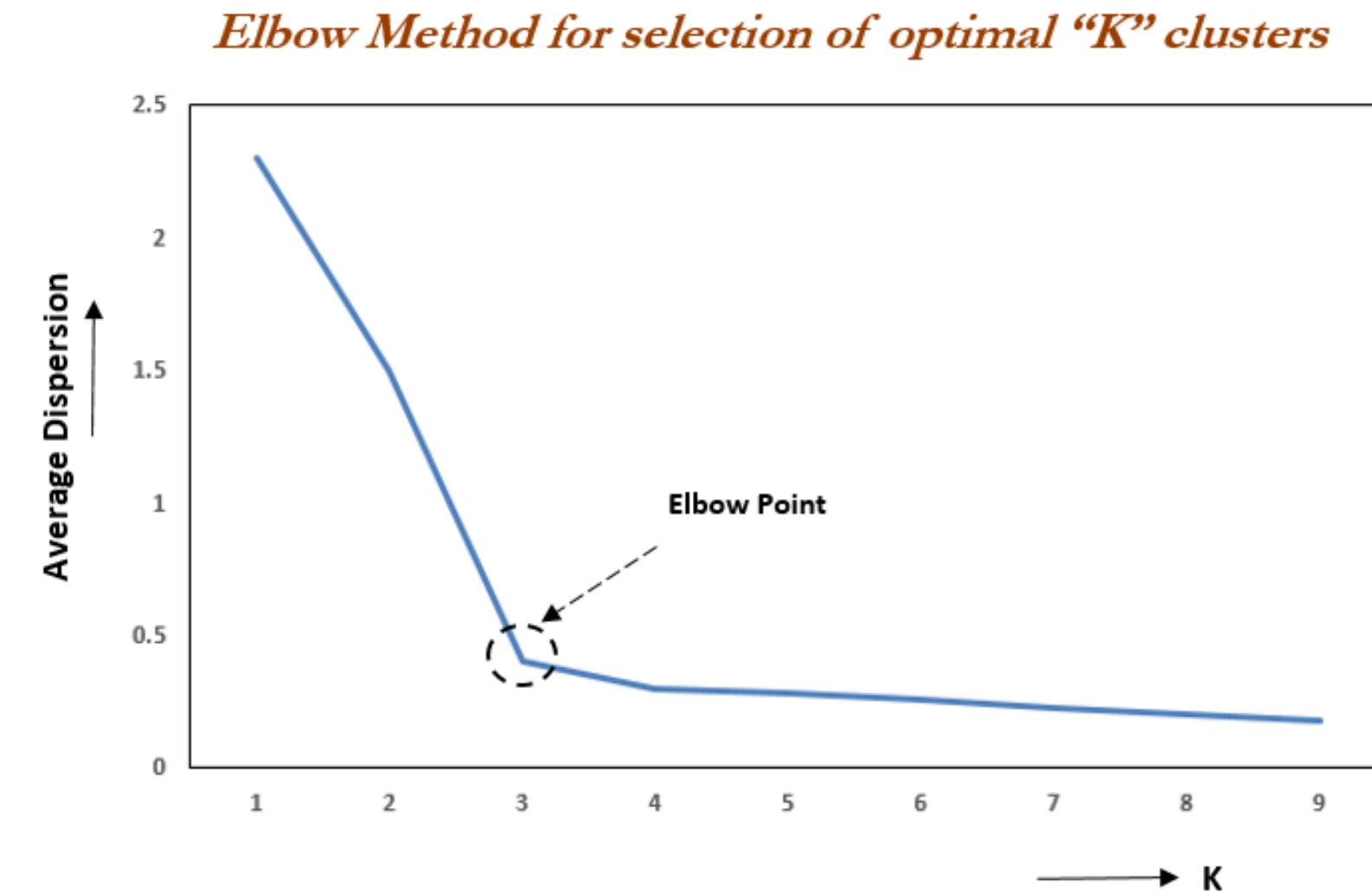
GIF source: shabal.in

ELBOW METHOD

Proprietary document of Indonesia AI 2023

Merupakan teknik yang digunakan untuk menentukan jumlah cluster yang optimal pada algoritma K-Means.

Elbow method mengacu pada kurva siku yang dihasilkan dari plot nilai SSE (sum of squared errors) atau inersia (inertia) terhadap jumlah cluster yang digunakan.



Indonesia AI

Image source: www.oreilly.com

SUM OF SQUARED ERRORS

Proprietary document of Indonesia AI 2023

SSE atau inersia adalah jumlah jarak kuadrat antara setiap titik data dalam kelompok dengan pusat kelompok mereka yang sesuai. SSE ini dapat digunakan sebagai ukuran kualitas cluster, di mana semakin kecil nilai SSE, semakin baik cluster yang dihasilkan.

Weights (x_i)	Mean	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
67	72.3	-5.3	28.09
86	72.3	13.7	187.7
62	72.3	-10.3	106.1
77	72.3	4.7	22.09
73	72.3	0.7	0.49
61	72.3	-11.3	127.7
80	72.3	7.7	59.29
75	72.3	2.7	7.29
69	72.3	-3.3	10.89
73	72.3	0.7	0.49

Indonesia AI

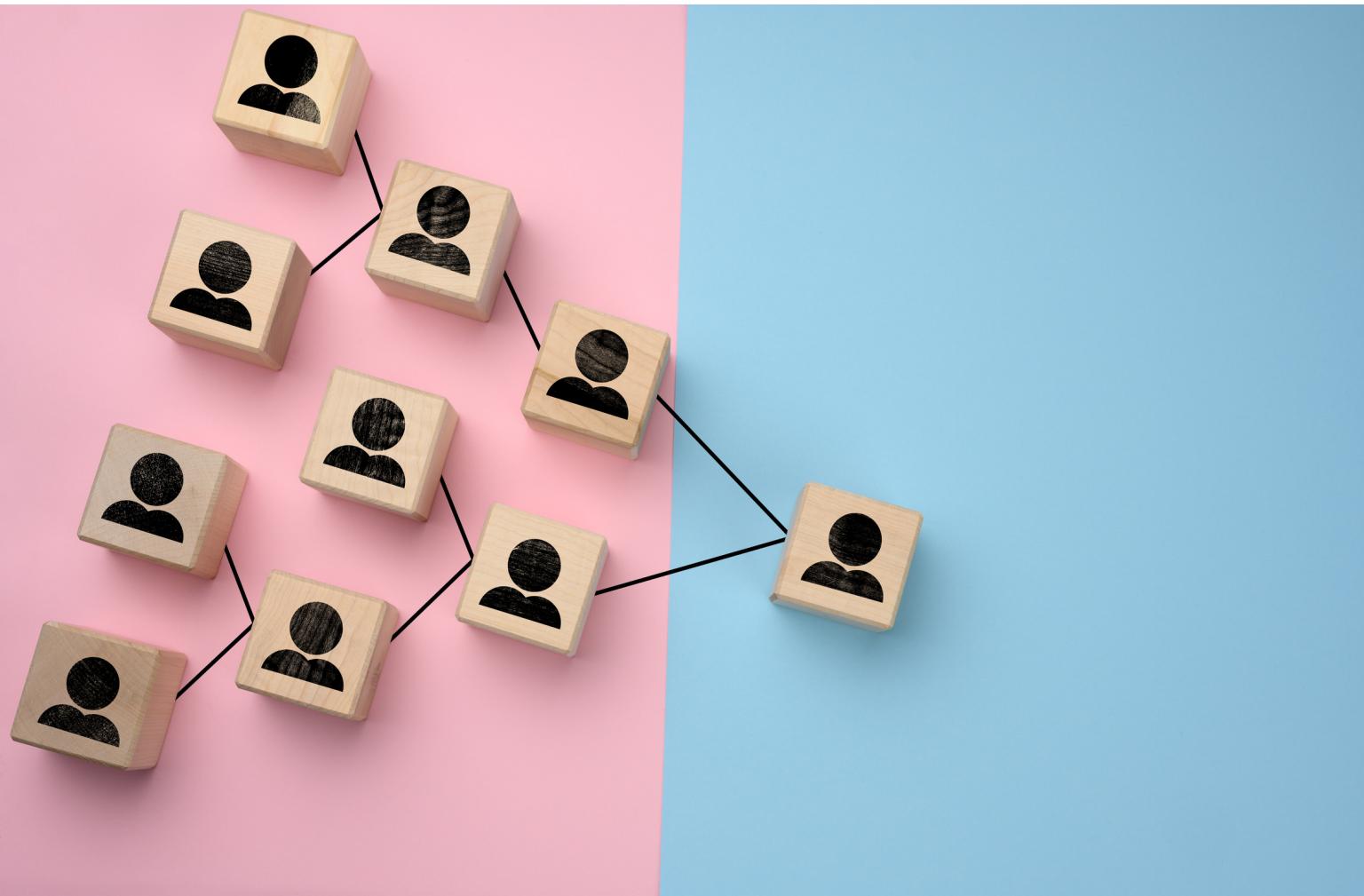
Image source: sixsigmastudyguide.com

Any question guys ~

Algoritma Hierarchical Clustering

DEFINISI

Proprietary document of Indonesia AI 2023

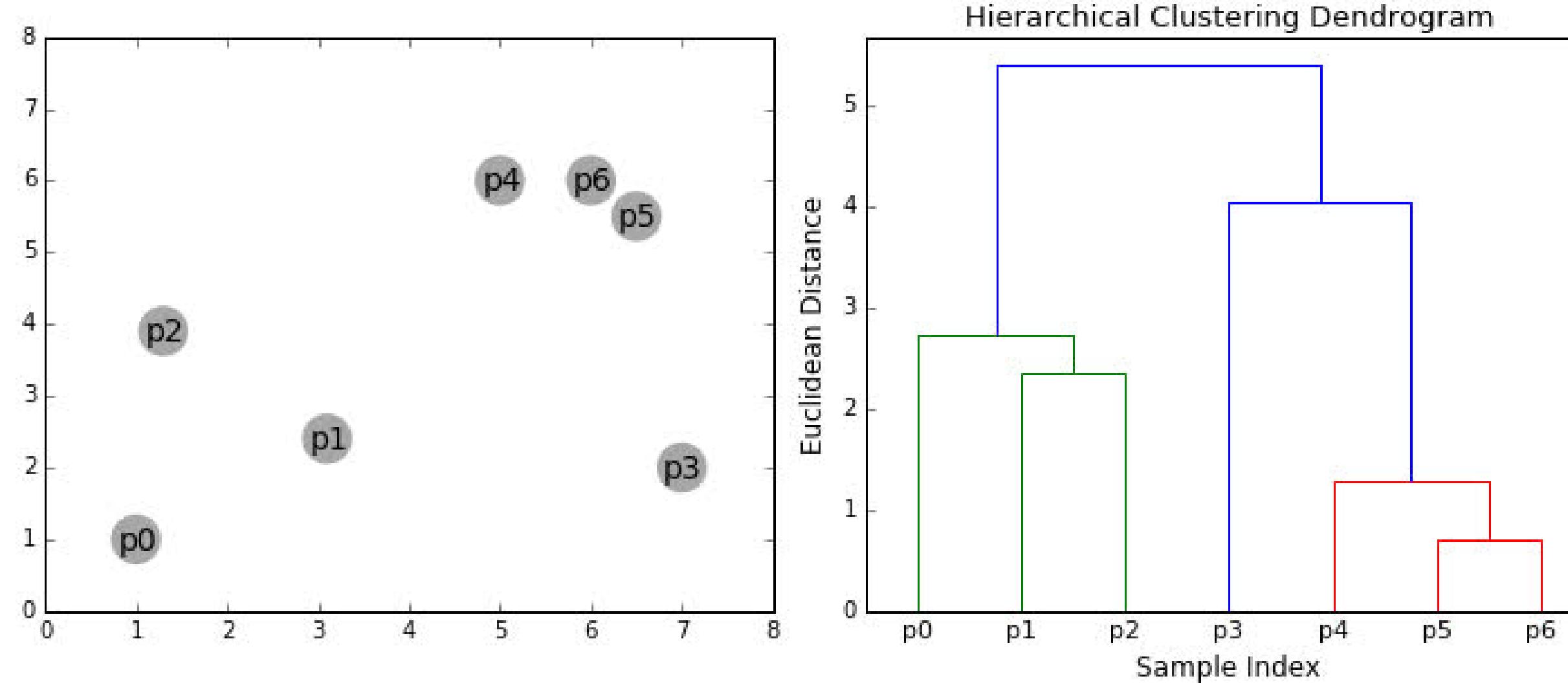


Hierarchical Clustering adalah algoritma Clustering dalam aplikasi Machine Learning yang mengelompokkan data ke dalam hierarki kelompok atau cluster berdasarkan tingkat kesamaan.

Dalam algoritma Hierarchical Clustering, kelompok yang dihasilkan disusun dalam struktur pohon atau dendogram, yang menunjukkan tingkat kesamaan (similarity) antar kelompok.

ILUSTRASI

Proprietary document of Indonesia AI 2023

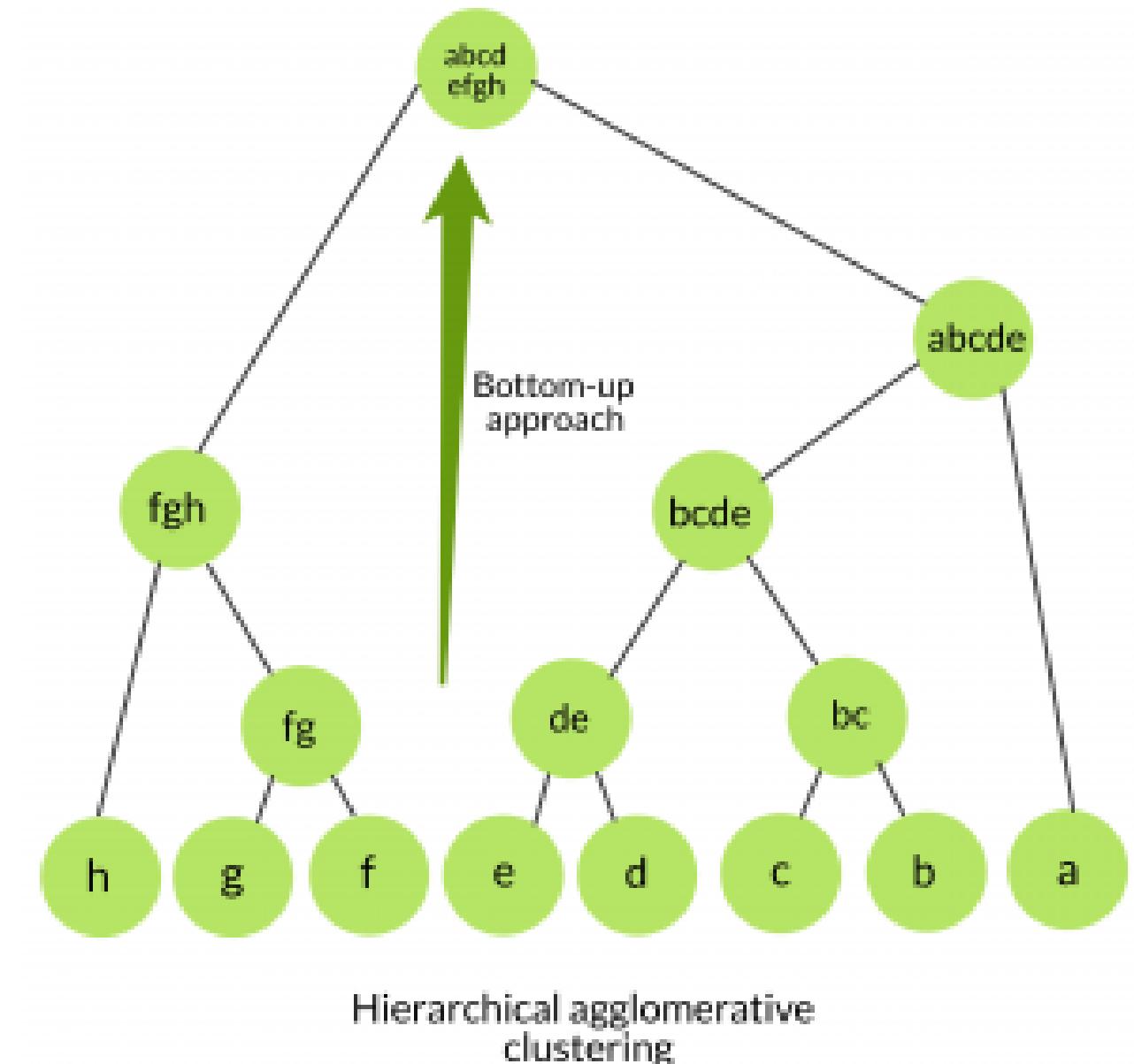


GIF source: David Sheehan

AGGLOMERATIVE HIERARCHICAL CLUSTERING

Proprietary document of Indonesia AI 2023

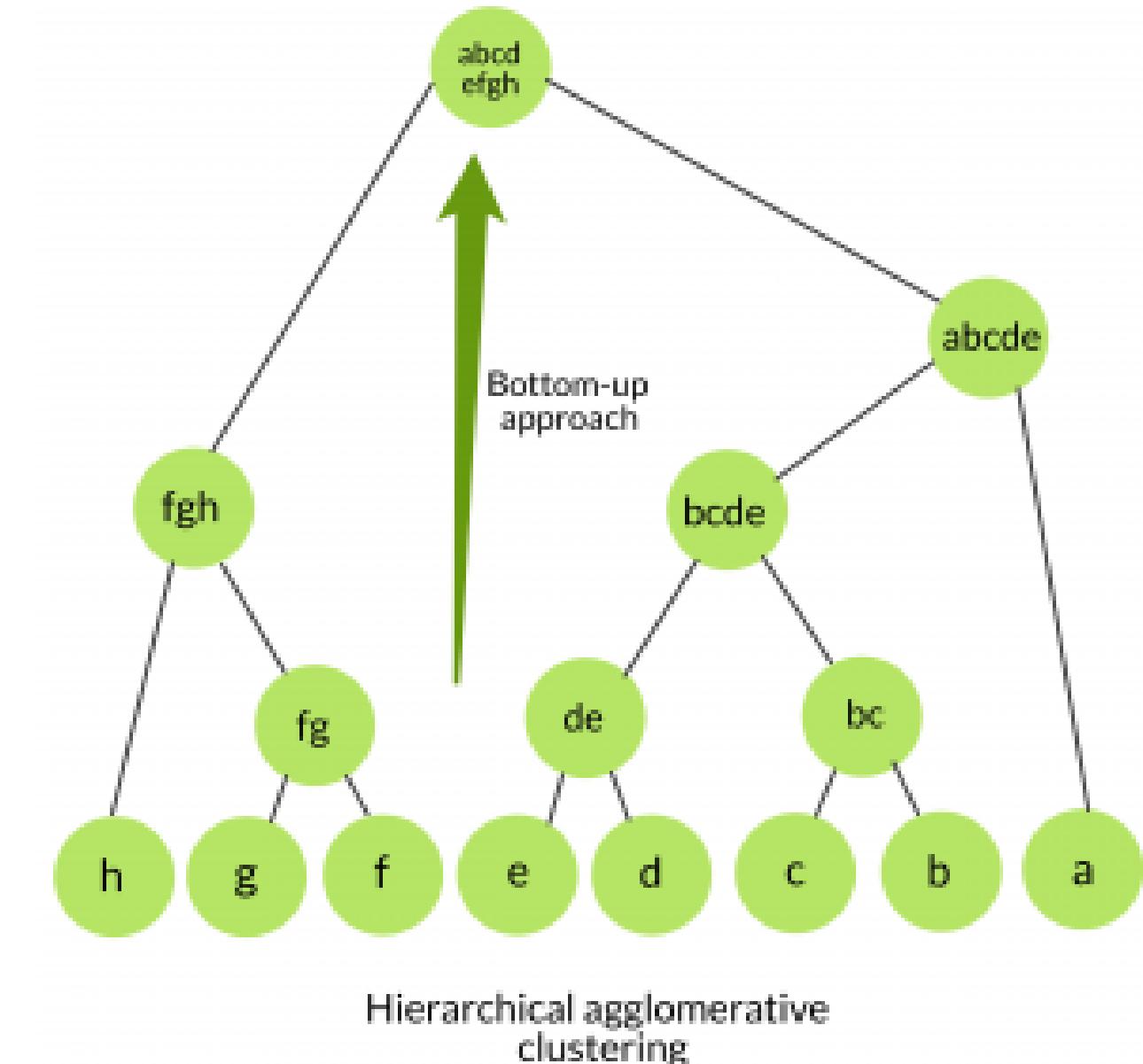
Pada Agglomerative Hierarchical Clustering, setiap titik data awalnya dianggap sebagai kelompok terpisah dan kemudian kelompok-kelompok tersebut digabungkan secara iteratif berdasarkan tingkat kesamaan antara titik data, hingga semua titik data digabungkan menjadi satu kelompok.



DIVISIVE HIERARCHICAL CLUSTERING

Proprietary document of Indonesia AI 2023

Pada Divisive Hierarchical Clustering, sebaliknya, semua titik data awalnya dikelompokkan bersama dan kemudian dipisahkan secara iteratif menjadi kelompok-kelompok yang semakin kecil berdasarkan tingkat perbedaan antar titik data.



KEUNTUNGAN

Proprietary document of Indonesia AI 2023



Hierarchical Clustering biasanya digunakan untuk data dengan jumlah sedang hingga kecil.

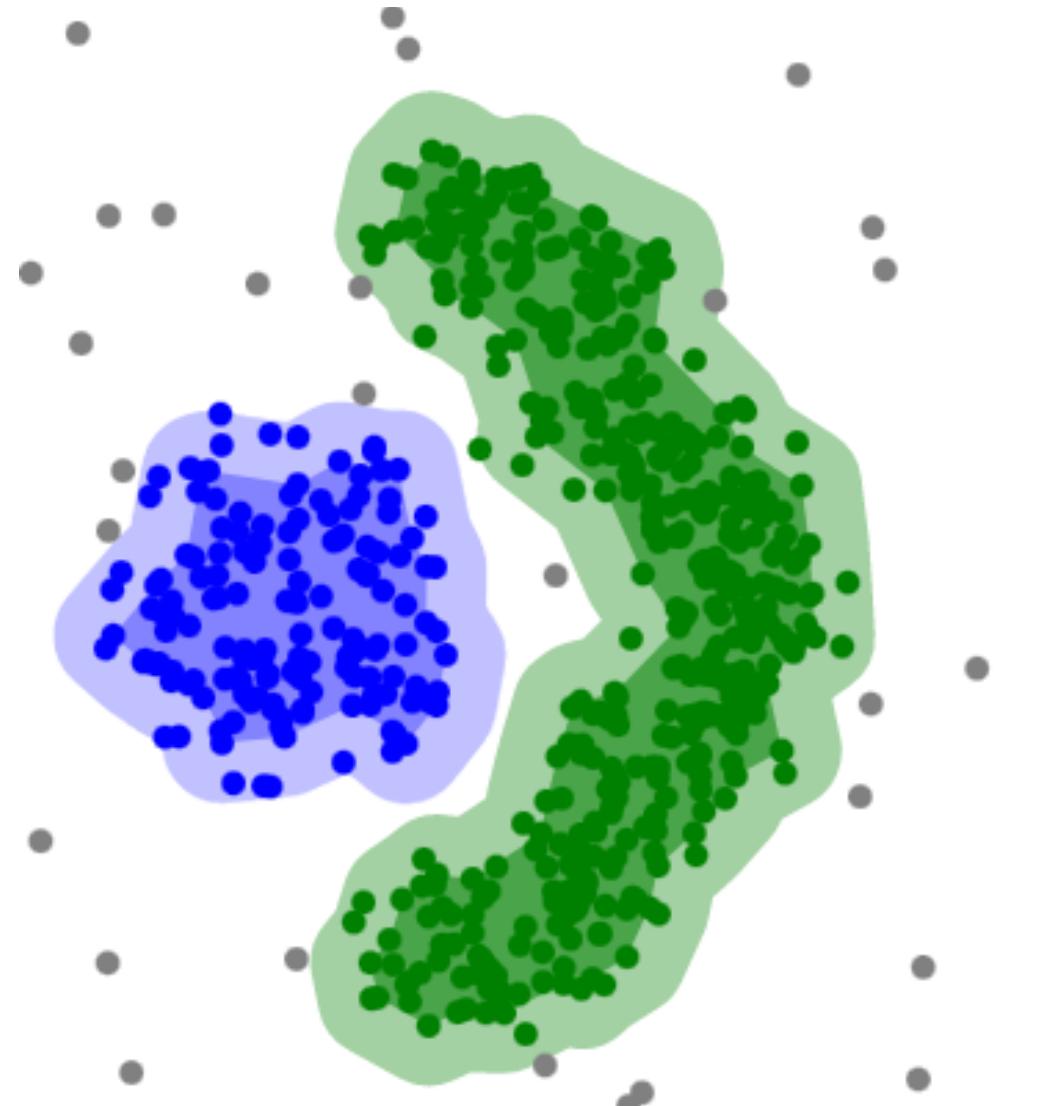
Keuntungan dari algoritma ini adalah bahwa hierarki cluster dapat **memberikan informasi tambahan tentang hubungan antara kelompok**, dan juga tidak perlu menentukan jumlah cluster secara eksplisit sebelumnya seperti K-Means Clustering misalnya.

Any question guys ~

— Algoritma DBSCAN

DEFINISI

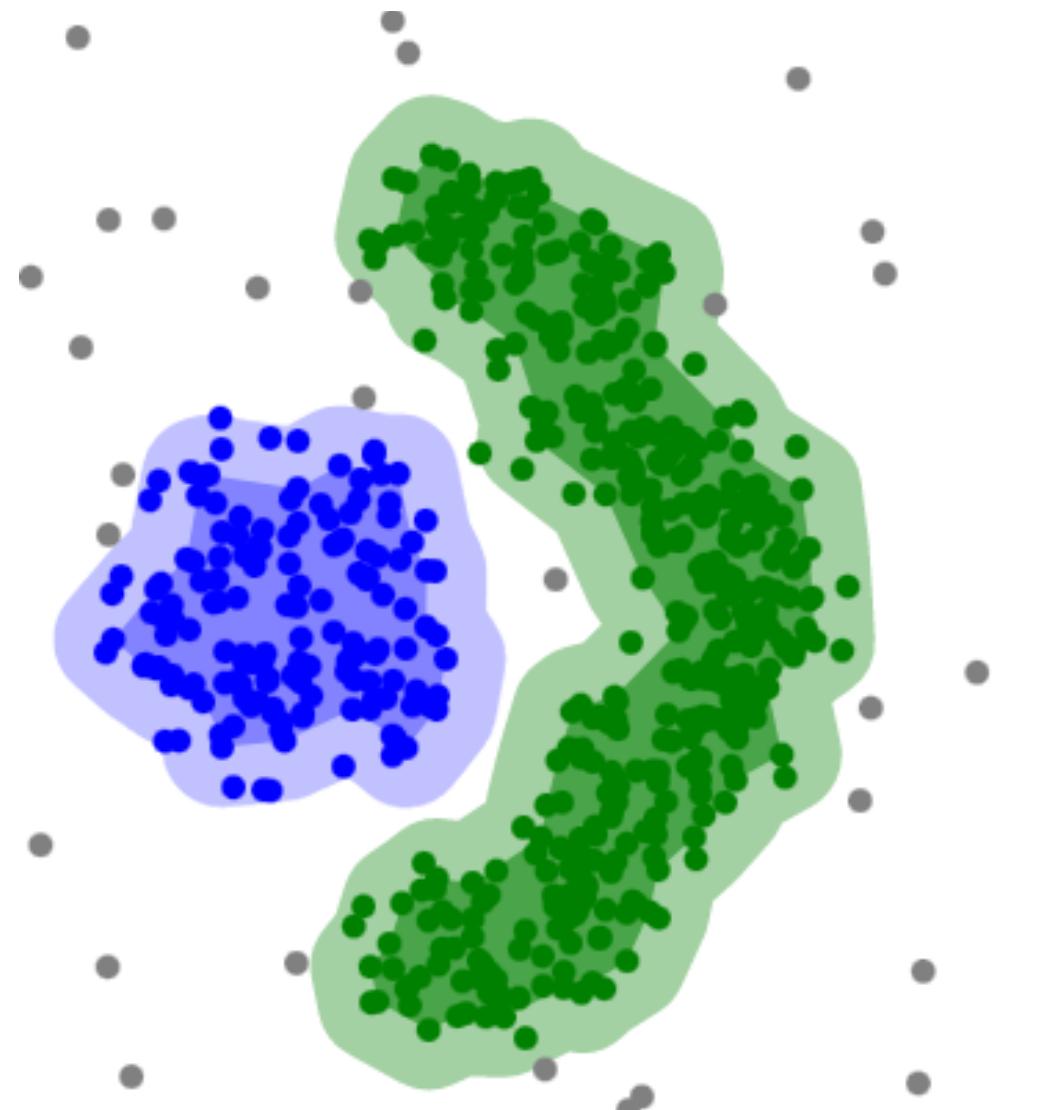
Proprietary document of Indonesia AI 2023



DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma clustering yang menggunakan kepadatan (density) data untuk mengelompokkan titik-titik data yang berdekatan menjadi cluster dan mengidentifikasi titik-titik data yang tidak tergabung dalam cluster sebagai noise.

CARA KERJA

Proprietary document of Indonesia AI 2023



Bagaimana algoritma DBSCAN bekerja?

1. Tentukan titik pusat (centroid) yang cukup dekat untuk membentuk sebuah cluster
2. Perluas cluster dengan menambahkan titik-titik data baru yang memiliki jarak yang cukup dekat dengan titik pusat (centroid)
3. Proses berlanjut sampai tidak ada lagi titik yang dapat ditambahkan ke dalam cluster

CARA KERJA

Proprietary document of Indonesia AI 2023

Titik-titik data yang tidak tergabung ke dalam cluster disebut sebagai **noise** atau **outlier**.

CARA KERJA

Proprietary document of Indonesia AI 2023

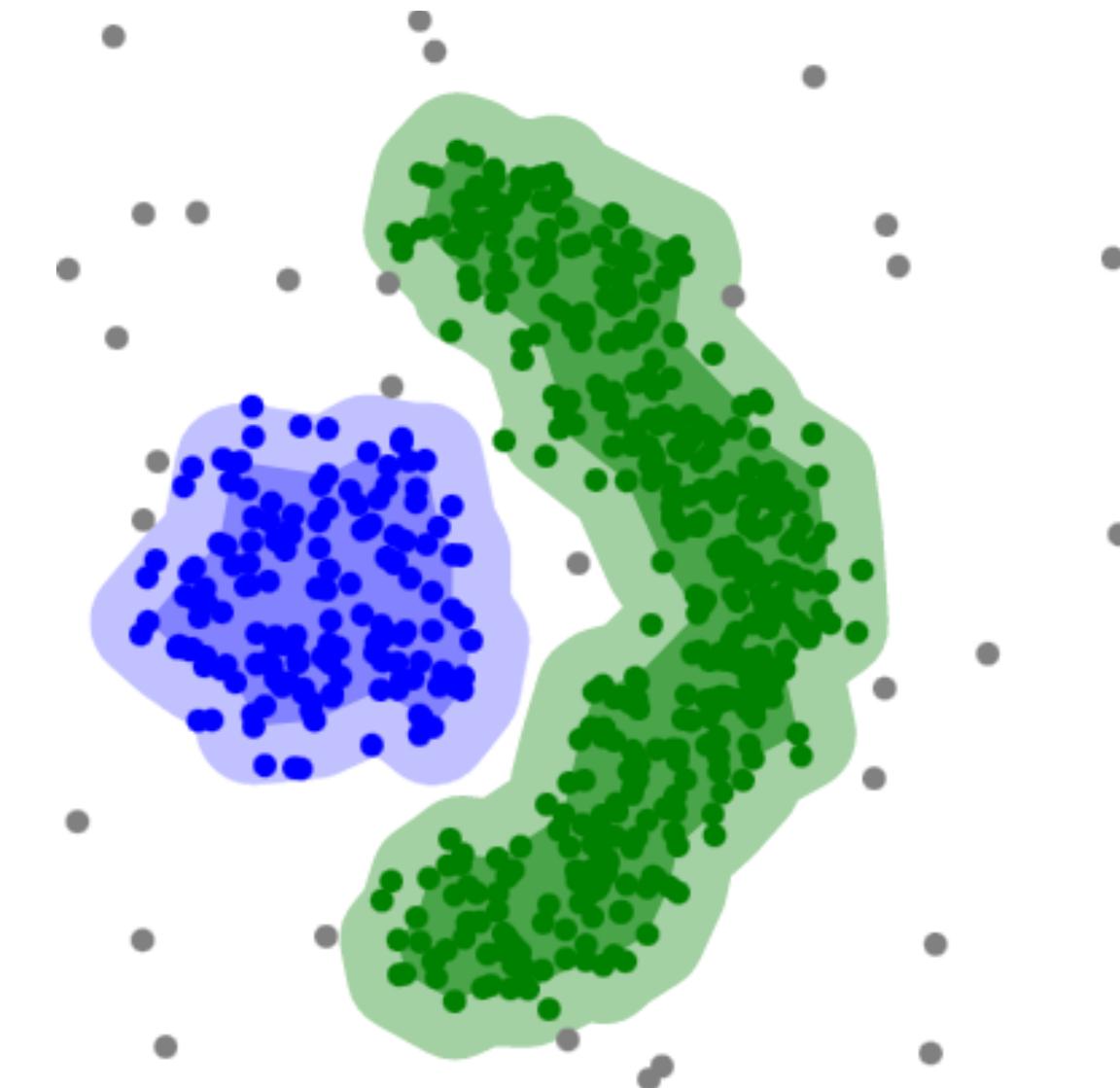


<https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/>

KEUNTUNGAN

Proprietary document of Indonesia AI 2023

Keuntungan dari algoritma DBSCAN adalah dapat menangani data yang kompleks dengan ukuran cluster yang bervariasi serta noise yang banyak, sehingga cocok untuk diterapkan pada data yang kompleks dan besar.

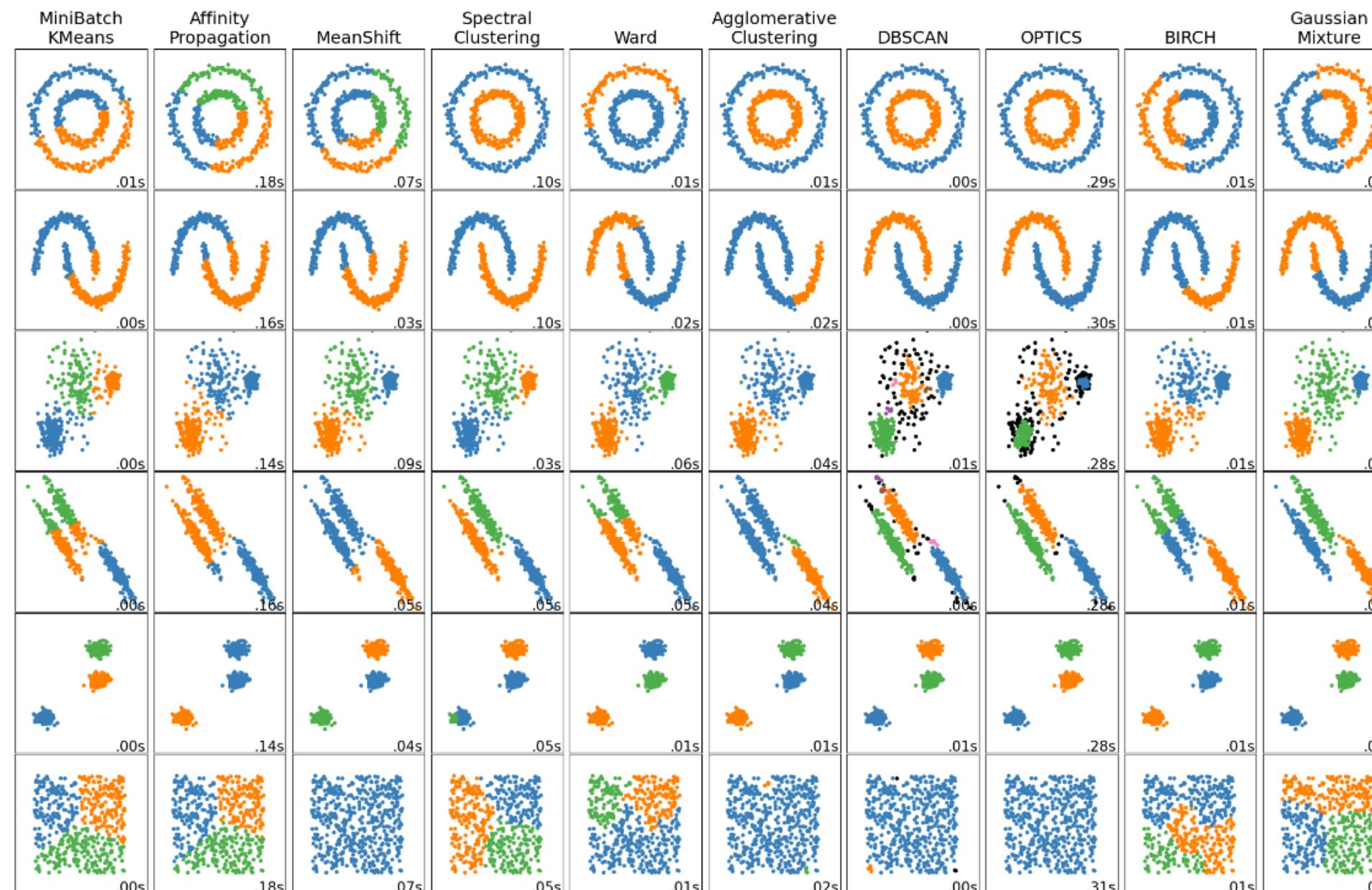


Indonesia AI

Image source: en.wikipedia.org

DAN MASIH BANYAK LAGI...

Proprietary document of Indonesia AI 2023



Any question guys ~

Terimakasih!