



**Indonesia AI**  
AI for Everyone, AI for Indonesia

**# 2023JAGO**  
**TEKNOLOGIAI**



## AI Career Bootcamp

# Clustering Algorithm

## Pembukaan

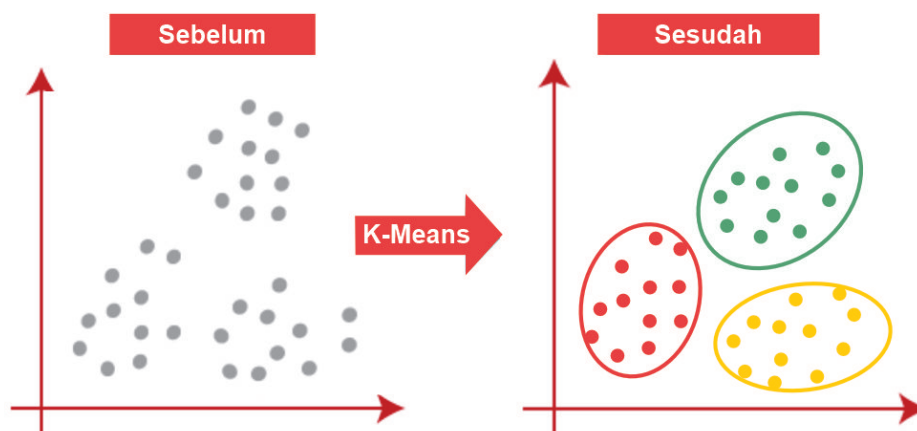
Guide Book ini memuat beberapa informasi-informasi utama yang akan disampaikan oleh mentor di program AI Career Bootcamp Kelas Computer Vision yang bisa dijadikan pegangan para students untuk mempersiapkan diri sebelum sesi Live Class berlangsung.

## Deskripsi

Clustering adalah salah satu algoritma Machine Learning yang digunakan untuk melakukan analisis data yang bertujuan untuk mengelompokkan data menjadi beberapa kelompok (cluster) yang memiliki karakteristik atau pola yang serupa. Algoritma Clustering merupakan teknik untuk melakukan kelompok (cluster) tersebut. Ada banyak sekali algoritma Clustering yang populer seperti K-Means, Hierarchical Clustering dan DBSCAN.

Algoritma Clustering terdiri dari beberapa tahap yakni.

- Tahap pertama adalah menentukan jumlah kelompok atau cluster yang diinginkan.
- Data yang akan dikelompokkan diolah dengan menggunakan metode tertentu, seperti K-Means atau Hierarchical Clustering. Metode K-Means membagi data menjadi kelompok berdasarkan jarak antara titik data dan centroid, sedangkan metode Hierarchical Clustering membangun struktur hirarki kelompok berdasarkan kemiripan data (similarity).
- Tahap selanjutnya adalah evaluasi hasil clustering. Evaluasi dapat dilakukan dengan menggunakan metrik yang berbeda, seperti misalnya Silhouette Coefficient. Metrik tersebut digunakan untuk mengukur seberapa baik hasil clustering yang dilakukan.



Contoh penggunaan algoritma Clustering adalah pada bidang bisnis. Misalnya, pada industri ritel, algoritma Clustering dapat digunakan untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka, sehingga dapat dilakukan strategi pemasaran yang lebih tepat sasaran. Selain itu, pada bidang kesehatan, algoritma Clustering dapat digunakan untuk mengelompokkan pasien berdasarkan jenis penyakit atau gejala yang serupa, sehingga dapat dilakukan penanganan yang lebih efektif dan tepat.

Algoritma Clustering termasuk dalam kelompok algoritma Unsupervised Learning karena tidak membutuhkan data dengan label untuk bisa bekerja.

## Algoritma K-Means

Algoritma K-Means Clustering bekerja dengan mengelompokkan data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan fitur antar data. Berikut adalah langkah-langkah algoritma K-Means Clustering secara detail:

- 1** Tentukan jumlah cluster yang diinginkan  
Langkah pertama dalam algoritma K-Means Clustering adalah menentukan jumlah cluster yang diinginkan. Jumlah cluster ini akan mempengaruhi hasil clustering akhir yang diperoleh.
- 2** Inisialisasi posisi centroid secara acak  
Setelah jumlah cluster ditentukan, langkah selanjutnya adalah menginisialisasi posisi centroid secara acak. Posisi centroid ini akan berperan sebagai pusat dari setiap cluster.
- 3** Kelompokkan data ke dalam cluster berdasarkan jarak terdekat  
Setelah posisi centroid ditentukan, data akan dikelompokkan ke dalam cluster berdasarkan jarak terdekat antara data dengan centroid terdekat. Jarak ini dihitung menggunakan suatu metrik jarak seperti Euclidean distance atau Manhattan distance.
- 4** Hitung posisi centroid baru berdasarkan rata-rata data pada setiap cluster  
Setelah data dikelompokkan ke dalam cluster, posisi centroid akan diperbarui berdasarkan rata-rata dari data yang termasuk ke dalam cluster tersebut. Posisi centroid baru ini akan menjadi pusat dari setiap cluster pada iterasi selanjutnya.

- 5 Ulangi langkah 3 dan 4 hingga tidak terjadi perubahan posisi centroid  
Langkah 3 dan 4 diulang kembali dengan mengelompokkan data ke dalam cluster baru berdasarkan posisi centroid yang telah diperbarui dan memperbarui posisi centroid hingga tidak terjadi perubahan posisi centroid dari iterasi sebelumnya.

Hasil akhir dari algoritma K-Means Clustering adalah data yang telah dikelompokkan ke dalam beberapa cluster. Setiap data akan termasuk ke dalam cluster yang memiliki centroid terdekat dengan data tersebut.

Dalam praktiknya, langkah-langkah di atas akan diulang beberapa kali dengan inisialisasi posisi centroid yang berbeda-beda untuk mencari hasil clustering yang paling optimal.

Berikut cuplikan singkat yang memberikan ilustrasi bagaimana Algoritma K-Means Clustering bekerja: <https://www.youtube.com/shorts/OrvltluxVQQ>

## Kelebihan & Kekurangan

Kelebihan algoritma clustering adalah:

- Algoritma Clustering adalah mampu mengelompokkan data secara otomatis tanpa memerlukan label pada data. Hal ini sangat berguna dalam situasi di mana label pada data sulit atau mahal untuk diperoleh. Selain itu,
- Algoritma Clustering juga dapat membantu dalam melakukan pemetaan data ke dalam kelompok-kelompok yang saling berhubungan, sehingga memudahkan dalam analisis data.

Adapun kekurangannya yaitu sebagai berikut:

- Kecenderungan dalam menghasilkan hasil clustering yang berbeda-beda pada setiap percobaan. Hal ini disebabkan oleh inisialisasi posisi centroid yang acak dan kemungkinan terjadinya local optimum. Oleh karena itu, perlu dilakukan beberapa percobaan dengan inisialisasi posisi centroid yang berbeda-beda untuk mendapatkan hasil clustering yang optimal.
- Algoritma Clustering juga cenderung tidak cocok untuk data dengan jumlah dimensi yang sangat besar atau data dengan banyak noise, karena dapat mempengaruhi kualitas hasil clustering.