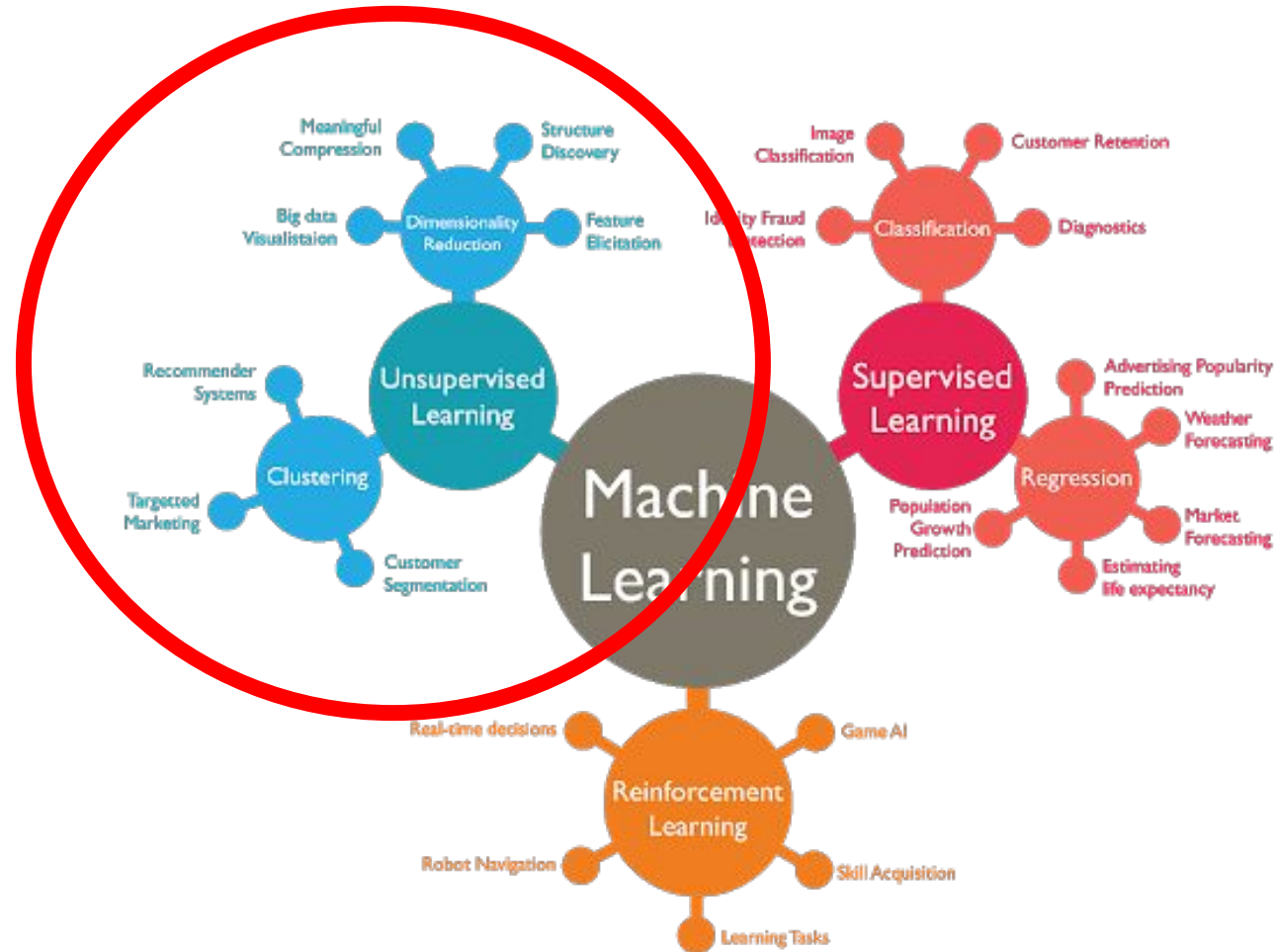


# Riset Perilaku Pelanggan (Unsupervised Learning)

# Machine Learning



# Unsupervised Learning

Unsupervised Learning merupakan tipe pembelajaran yang mirip seperti proses siswa belajar sendiri dengan melihat kesamaan yang ada. Pada unsupervised learning, mesin tidak diberikan informasi berupa label apa yang akan diprediksi.



# Unsupervised Learning

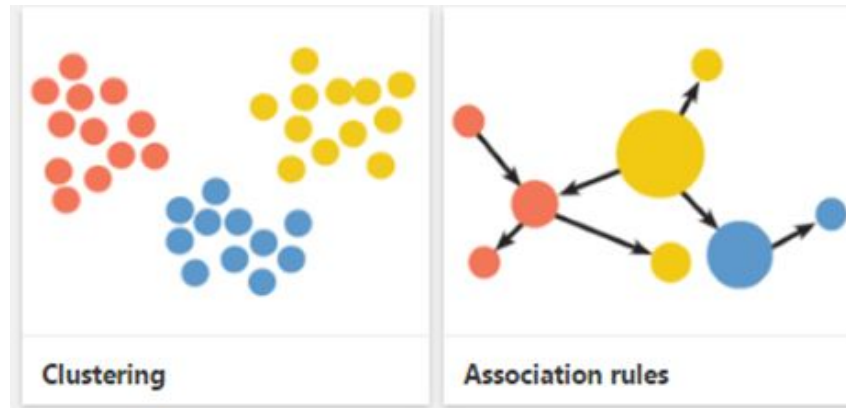


Bentuk	Warna	Tekstur	Buah
Lonjong	Kuning	Lembut	1
Lonjong	Kuning	Lembut	1
Bulat	Kuning	Keras	2

# Unsupervised Learning

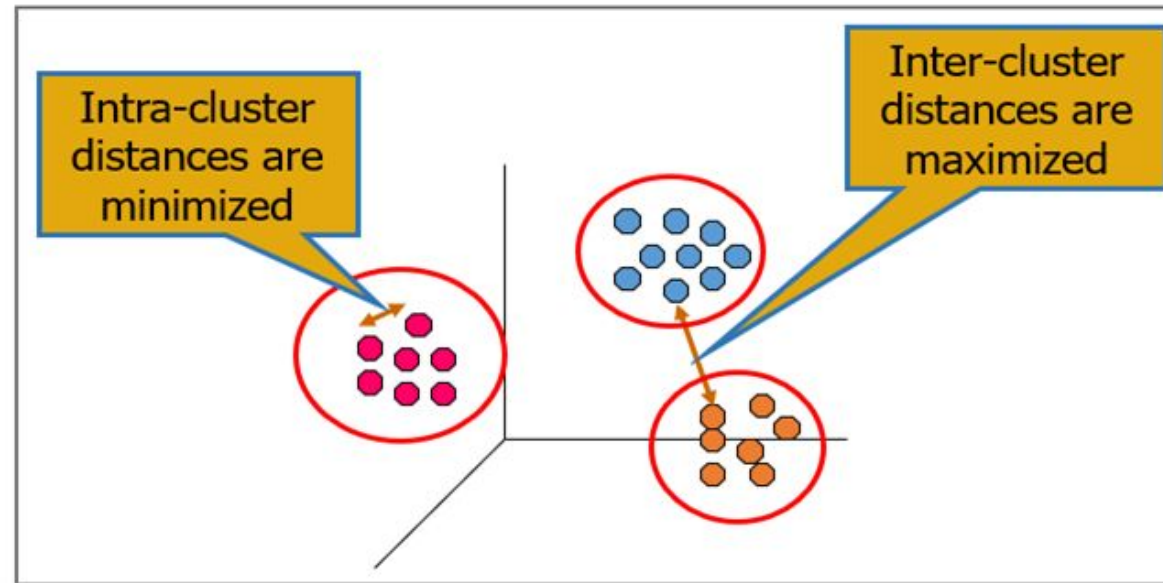
Unsupervised Learning ini umum digunakan untuk menyelesaikan 2 tipe masalah:

- **Clustering:** Mengelompokkan data berdasarkan persamaan karakteristik
- **Association:** Menemukan asosiasi di antara item dalam data transaksi yang besar



# Clustering

Menemukan kelompok objek yang tepat sehingga objek di dalamnya akan sama (berhubungan) satu sama lain dan berbeda (tidak berhubungan) dengan objek di kelompok lain.



# Aplikasi Clustering

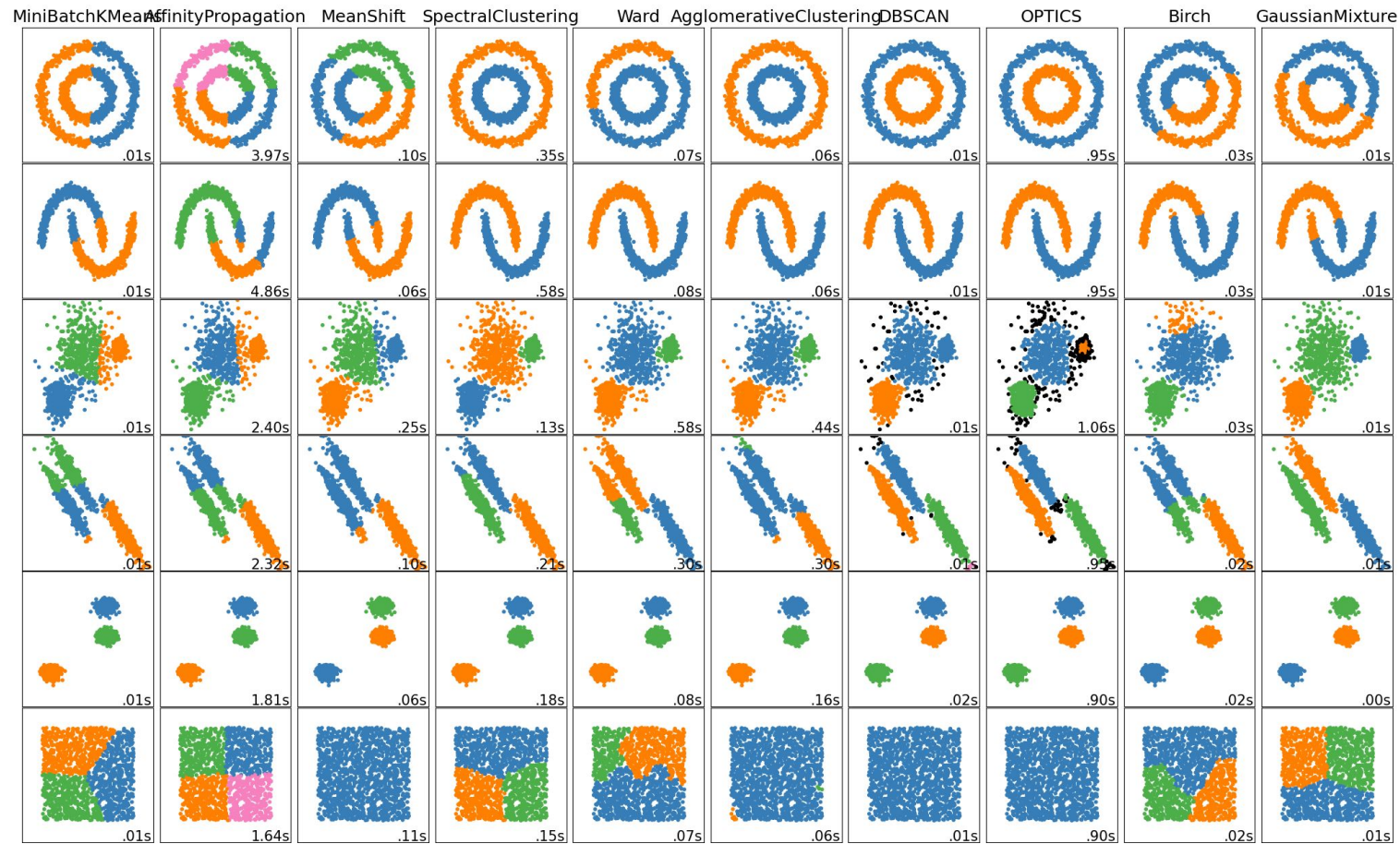
- Segmentasi Pelanggan
- Pengoptimalan Toko Pengiriman
- Klasifikasi Dokumen

# Teknik Clustering

- Partitional Clustering
- Hierarchical Clustering
- Density Based Clustering

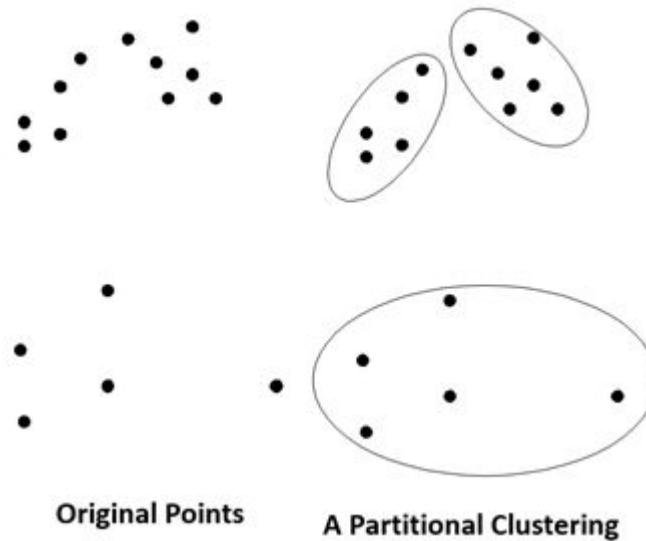


# Teknik Clustering Scikit-Learn



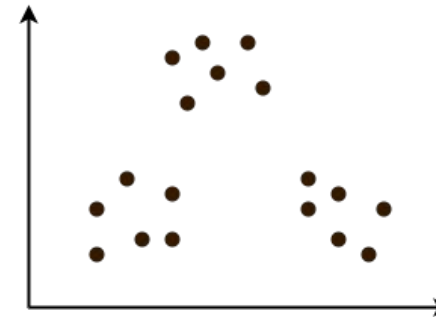
# Partitional Clustering

Partitional clustering (or partitioning clustering) are clustering methods used to classify observations, within a data set, into multiple groups based on their similarity.

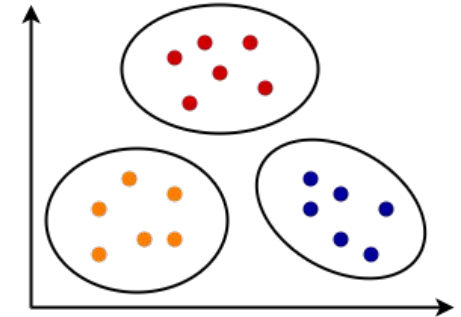


# K-Means Clustering

- Algoritma yang paling sederhana
- Setiap klaster dikaitkan dengan pusat/centroid
- Setiap titik ditempatkan ke klaster dengan titik pusat terdekat
- Jumlah klaster,  $K$ , harus ditentukan

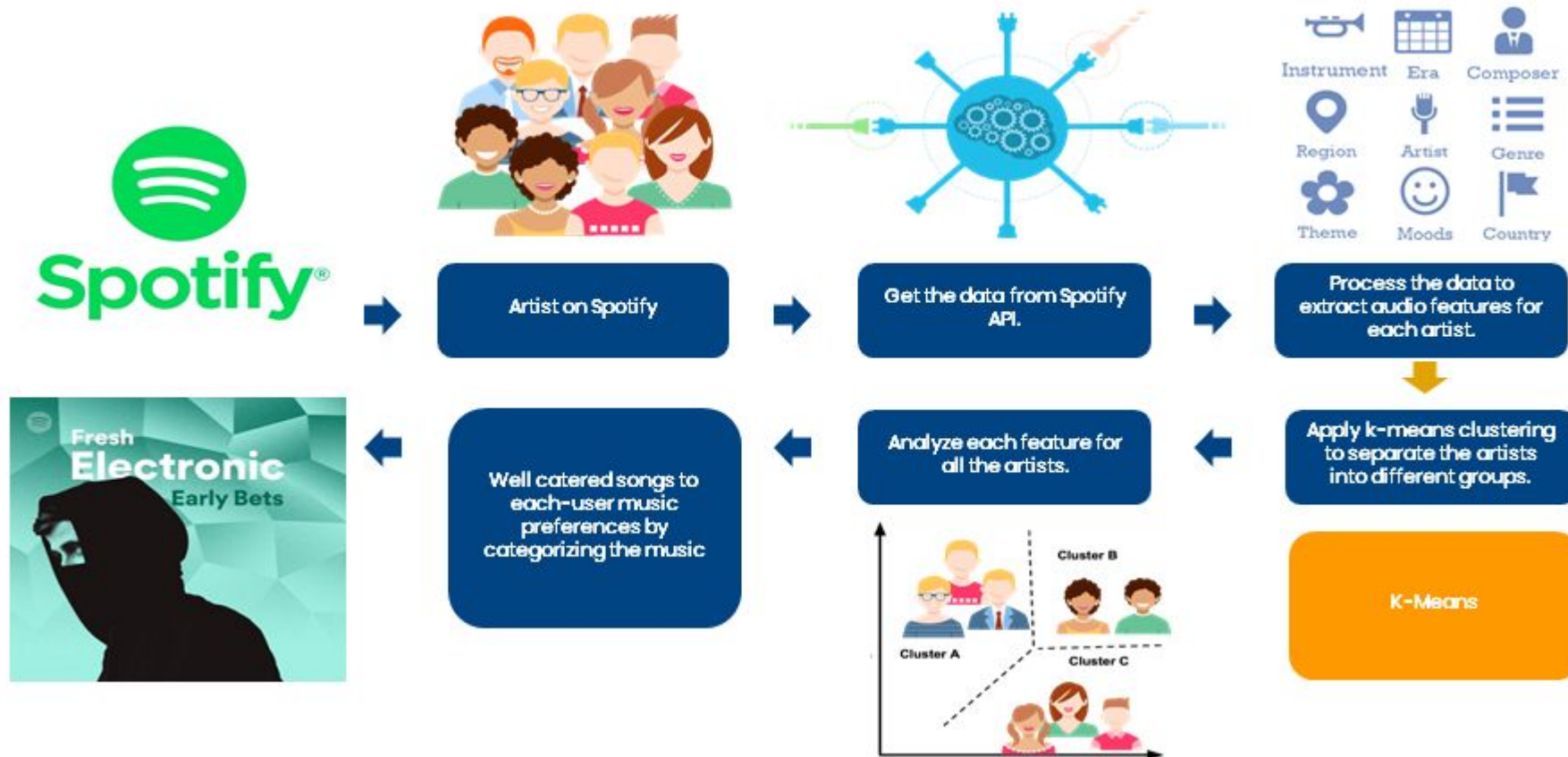


Before K-Means



After K-Means

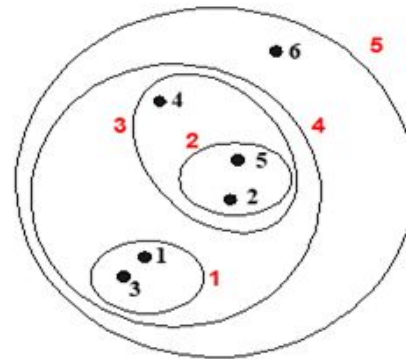
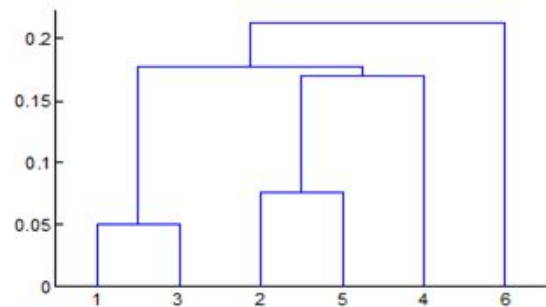
# K-Means in Spotify



# Hierarchical Clustering

Menghasilkan klaster yang diatur sebagai pohon hierarki.

Dapat divisualisasikan dengan dendrogram, seperti diagram pohon yang mencatat urutan penggabungan atau pemisahan.



# Jenis Hierarchical Clustering

## **Agglomerative (Bottom-Up):**

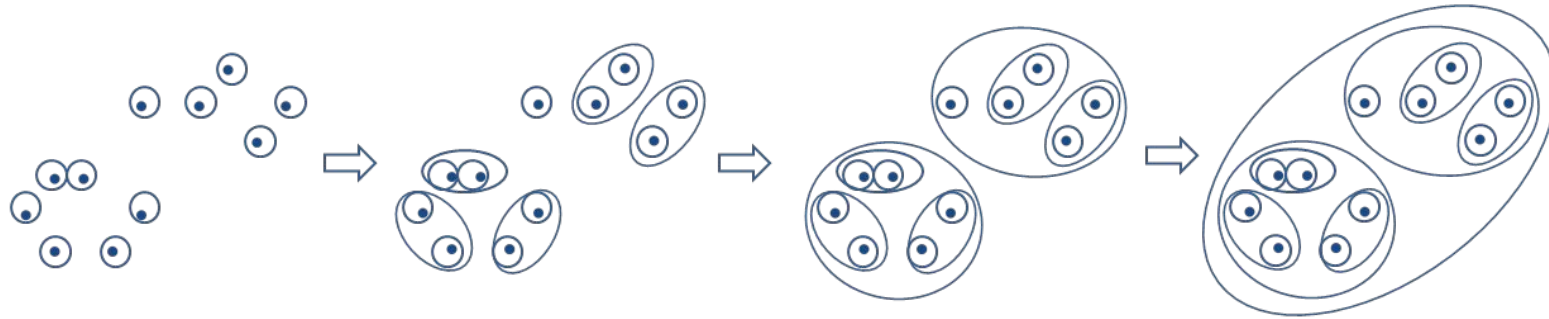
- Dimulai dengan titik sebagai kelompok individu
- Di setiap tahap, gabungkan klaster terdekat hingga hanya tersisa satu klaster (k klaster)

## **Divisive (Top-Down) :**

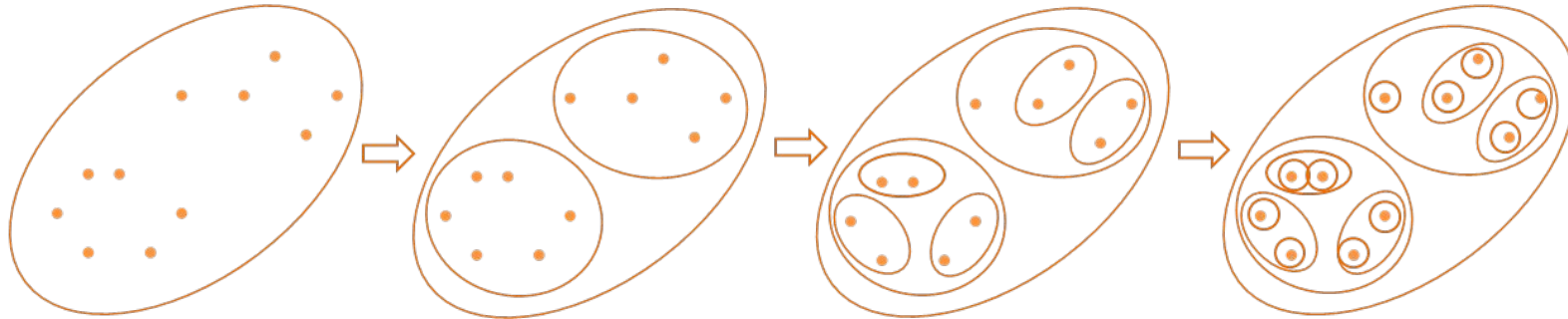
- Dimulai dengan satu klaster
- Setiap tahap, pisahkan satu per satu klaster hingga setiap klaster hanya memiliki satu titik ( hanya ada k klaster)

# Jenis Hierarchical Clustering

Agglomerative Hierarchical Clustering



Divisive Hierarchical Clustering



Sumber: <https://towardsdatascience.com/https-towardsdatascience-com-hierarchical-clustering-6f3c98c9d0ca>

# Kelebihan Hierarchical Clustering

- Tidak perlu mengasumsikan jumlah klaster tertentu. Jumlah klaster dapat diperoleh dengan 'memotong' bagian dendogram pada tingkat yang tepat.
- Memungkinkan bersesuaian dengan taksonomi.



# Association Rule

- Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item.
- Contoh: Mengetahui seberapa berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu.
- Sering disebut dengan istilah market basket analysis.



Sumber: [http://www.saedsayad.com/association\\_rules.htm](http://www.saedsayad.com/association_rules.htm)

# Association Rule

<i>TID</i>	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}

# Association Rule

<i>TID</i>	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}

## Contoh Aturan:

- {Milk,Diaper} → {Beer}
- {Milk,Beer} → {Diaper}
- {Diaper,Beer} → {Milk}

# Association Rule Metric

- **Support:** Bagian transaksi yang mengandung X dan Y
- **Confidence:** Mengukur seberapa sering item di Y muncul dalam transaksi yang mengandung X
- **Lift:** Rasio kemunculan suatu rules dibandingkan dengan kemunculan masing-masing item

# Association Rule

Rule:  $X \Rightarrow Y$

$$Support = \frac{freq(X, Y)}{N}$$

$$Confidence = \frac{freq(X, Y)}{freq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$



Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B \& C \Rightarrow D$	1/5	1/3	5/9

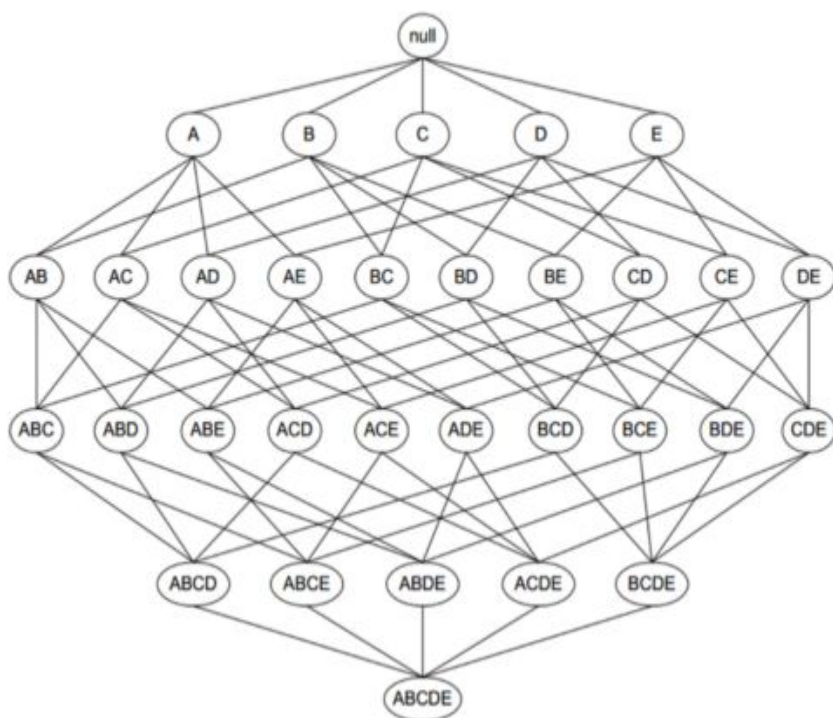
Sumber: [http://www.saedsayad.com/association\\_rules.htm](http://www.saedsayad.com/association_rules.htm)

# Apriori

- Pembuatan rule masih mahal secara komputasi
- Algoritma Apriori mengurangi Jumlah Kandidat rule
- Jika suatu pasangan item sering muncul, maka subsetnya juga harus demikian, dan sebaliknya

# Apriori

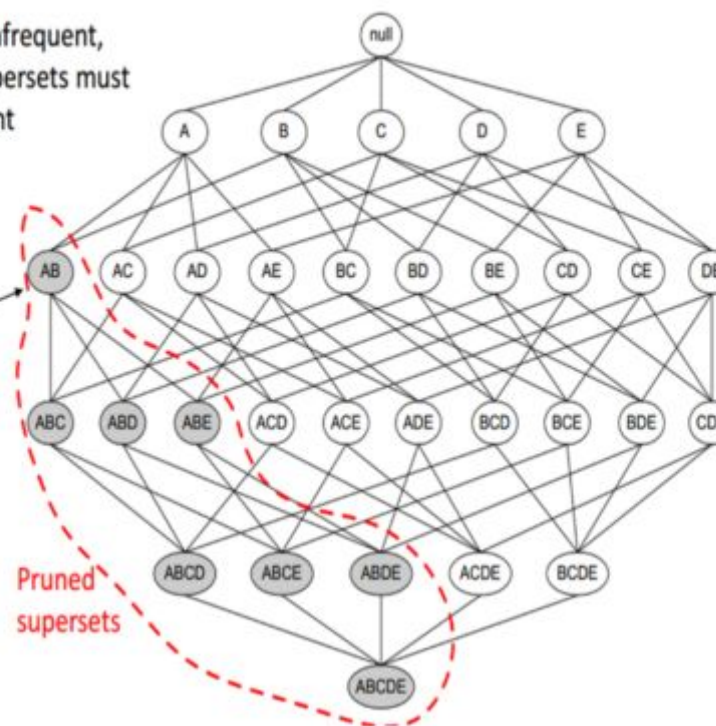
The combinations of 5 items



The Apriori Algorithm

If an itemset is infrequent,  
then all of its supersets must  
also be infrequent

Found to be  
Infrequent



<https://communities.sas.com/t5/SAS-Communities-Library/Association-Discovery-the-Apriori-algorithm/ta-p/569042>



# Praktek