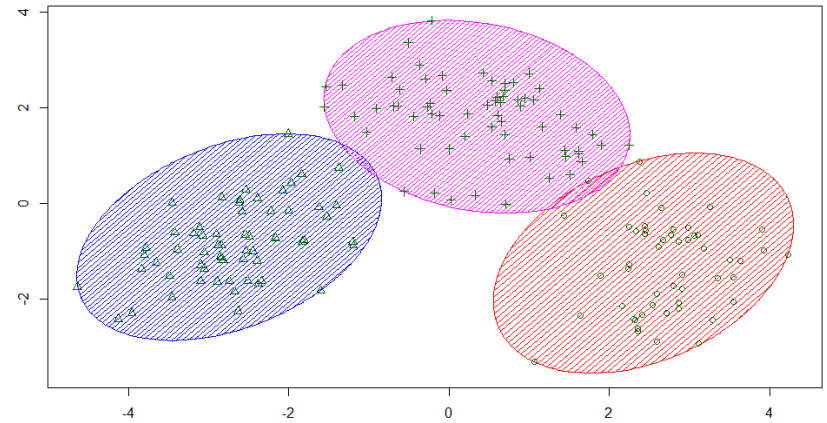


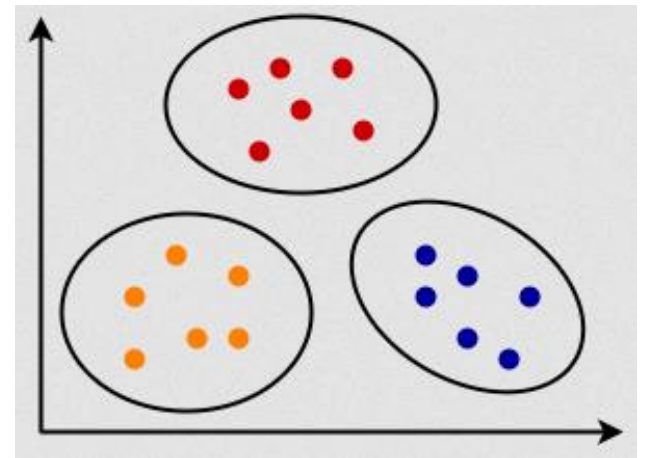
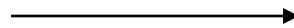
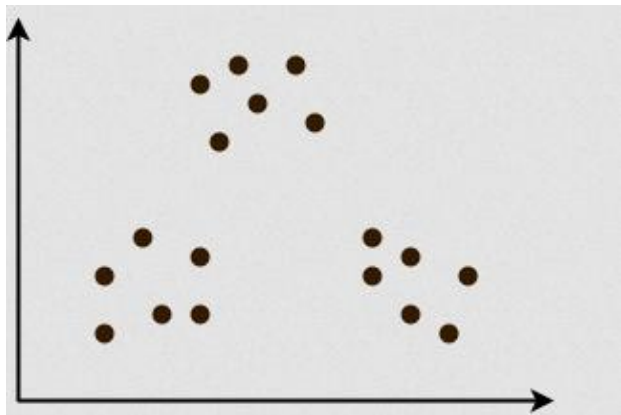
CLUSTERING

K-Means

Seno Alrianto



Cluster analysis adalah pengamatan kelompok berdasarkan fitur (atribut/variable) yang tersedia pada data tanpa adanya label kelompok dari setiap datanya untuk menghasilkan label pada masing-masing data yang sesuai.

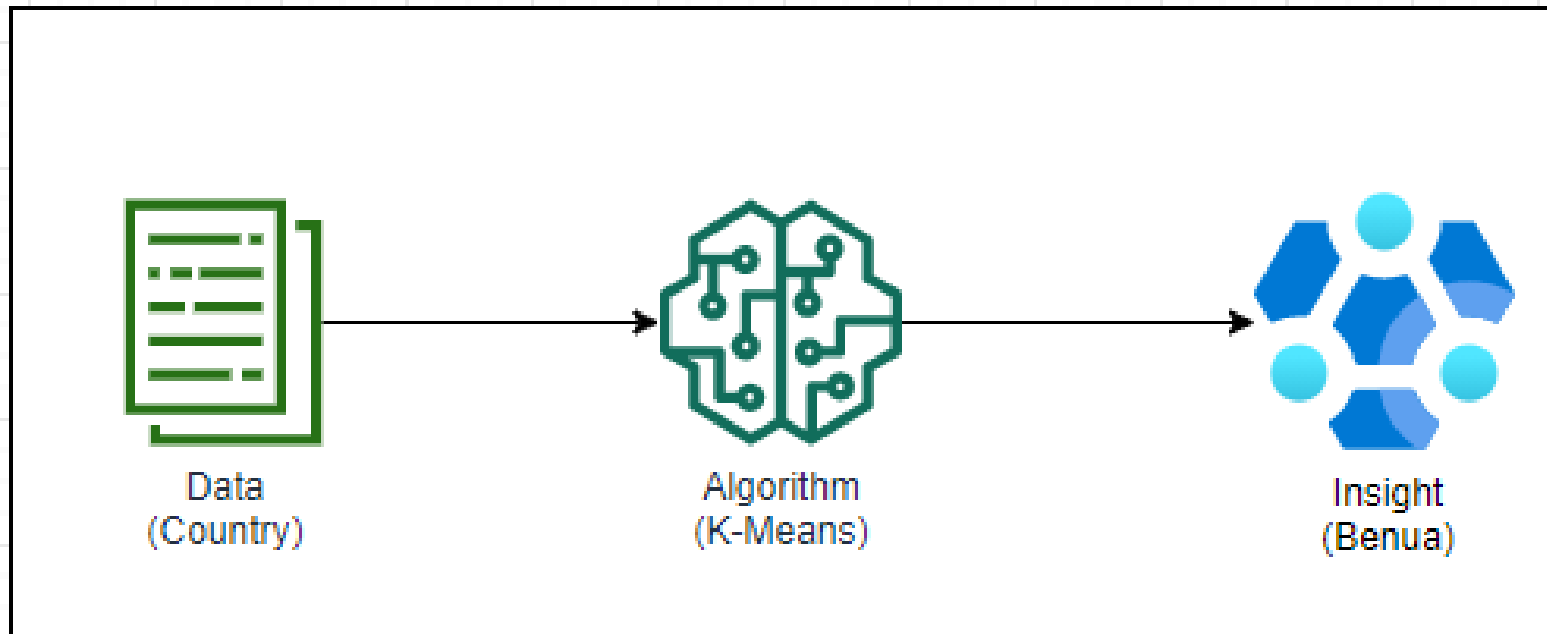


CONTOH



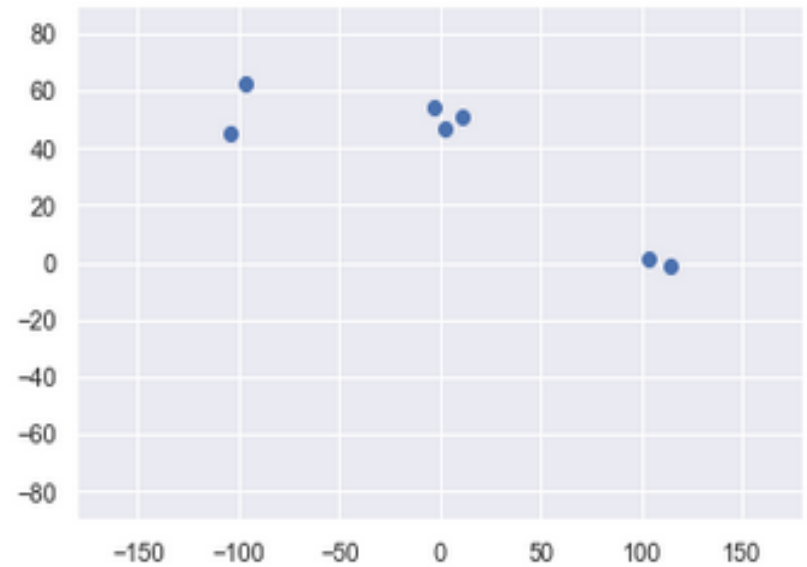
Pengelompokan beberapa negara di dunia untuk dibuat menjadi 3 klaster (dalam contoh ini menggunakan benua) untuk memudahkan label pada setiap data sebagai wawasan pengguna.

Alur proses




Data

Country	Latitude	Longitude	Language
Indonesia	-0.78	113.92	Indonesia
Singapore	1.29	103.85	English
UK	54.01	-2.53	English
Germany	51.15	10.40	German
USA	44.97	-103.77	English
France	46.75	2.40	French
Canada	62.40	-96.80	English



Seleksi Fitur :

digunakan latitude dan longitude dari setiap negara untuk mengetahui klaster berdasarkan data numerik yang tersedia.



Latitude	Longitude
-0.78	113.92
1.29	103.85
54.01	-2.53
51.15	10.40
44.97	-103.77
46.75	2.40
62.40	-96.80

K-Means

number of clusters number of cases centroid for cluster j

objective function $\leftarrow J = \sum_{j=1}^k \sum_{i=1}^n \underbrace{\|x_i^{(j)} - c_j\|}_\text{Distance function}^2$

case i

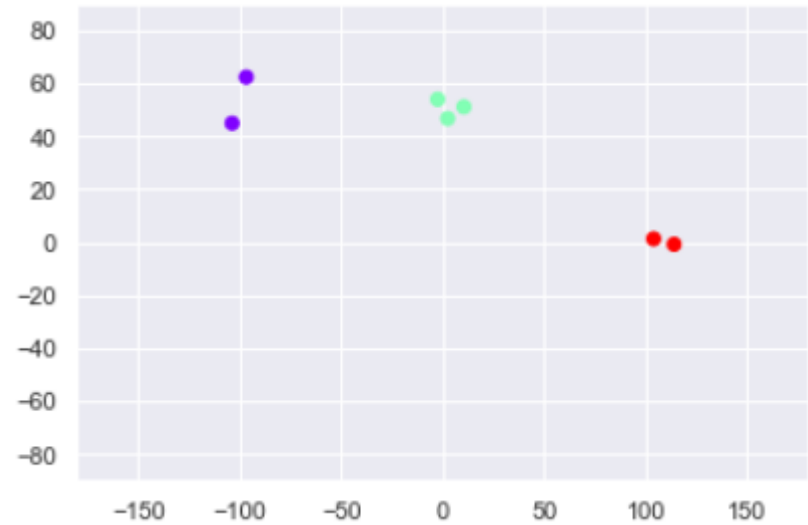
centroid for cluster j

Distance function

The diagram shows the objective function $J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$. Annotations include: 'number of clusters' pointing to k ; 'number of cases' pointing to n ; 'case i ' pointing to $x_i^{(j)}$; 'centroid for cluster j ' pointing to c_j ; 'Distance function' pointing to the norm $\|x_i^{(j)} - c_j\|$; and 'objective function' pointing to J . A blue arrow points from the 'objective function' label to J . A blue bracket under the norm is labeled 'Distance function'. A blue arrow points from 'centroid for cluster j ' to c_j . A blue arrow points from 'case i ' to $x_i^{(j)}$. A blue arrow points from 'number of clusters' to k . A blue arrow points from 'number of cases' to n .

Result

Country	Latitude	Longitude	Language	Cluster
Indonesia	-0.78	113.92	Indonesia	2
Singapore	1.29	103.85	English	2
UK	54.01	-2.53	English	1
Germany	51.15	10.40	German	1
USA	44.97	-103.77	English	0
France	46.75	2.40	French	1
Canada	62.40	-96.80	English	0



Insight

Country	Latitude	Longitude	Language	Cluster
Indonesia	-0.78	113.92	Indonesia	Asia
Singapore	1.29	103.85	English	Asia
UK	54.01	-2.53	English	Europe
Germany	51.15	10.40	German	Europe
USA	44.97	-103.77	English	America
France	46.75	2.40	French	Europe
Canada	62.40	-96.80	English	America



TERIMA KASIH



<https://www.linkedin.com/in/seno-alrianto-aa2615166/>



<https://salrianto.medium.com/>



<https://github.com/senoal>



https://www.saedsayad.com/clustering_kmeans.htm