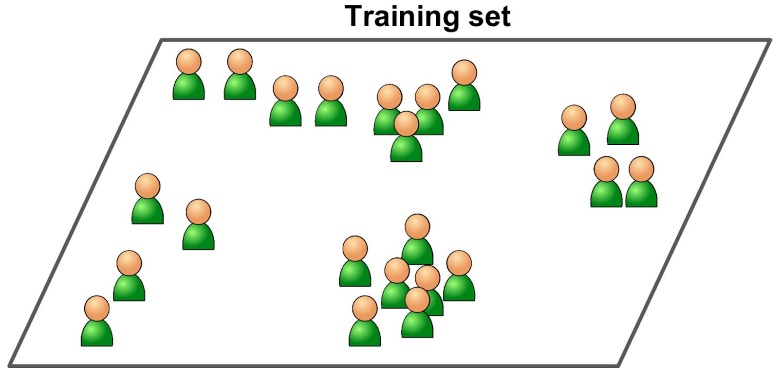
Klaster (cluster) adalah sebuah grup yang memiliki kemiripan tertentu. Pengklasteran adalah sebuah metode machine learning unsupervised untuk mengelompokkan objek-objek yang memiliki kemiripan, ke dalam sebuah klaster. Karena termasuk kategori unsupervised, maka dataset yang digunakan model clustering tidak memiliki label.

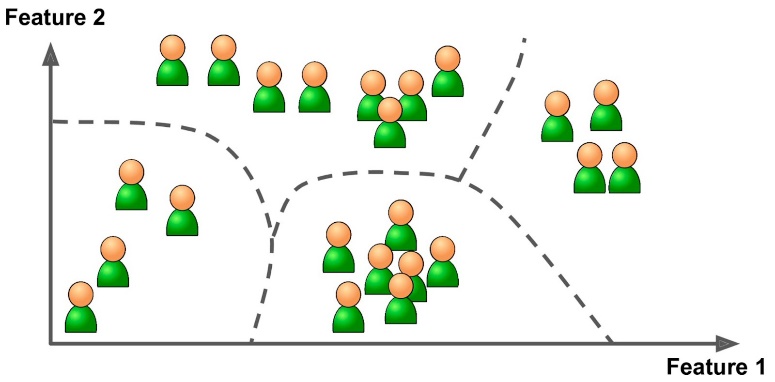
Menurut Andriy Burkov dalam buku The Hundred Page Machine Learning Book [8], clustering atau pengklasteran adalah sebuah metode untuk memberi label pada data tanpa bantuan manusia. Bagaimana proses pemberian label ini?

Data yang memiliki kemiripan akan dikelompokkan, lalu setiap data pada kelompok yang sama akan diberikan label yang sama. Berbeda kan, dengan model klasifikasi dan regresi di mana setiap data memiliki label yang ditulis oleh manusia.

Contohnya adalah ketika kita memiliki data pengunjung web toko online kita seperti gambar di bawah. Kemudian kita ingin mengembangkan sebuah model yang bisa mengelompokkan pengunjung yang memiliki kemiripan. Misalnya, diketahui bahwa 80% pengunjung toko online Anda adalah perempuan, sementara 20% nya adalah laki-laki. 60% dari pengunjung perempuan mengunjungi toko online Anda pada hari kerja, sementara sisanya berkunjung pada akhir minggu. Contoh lain, 40% pengunjung toko online Anda berasal dari Pulau Jawa, 55% berasal dari pulau lain di seluruh Indonesia, dan 5% sisanya berasal dari luar negeri. Tujuan pengelompokkan kemiripan ini adalah agar kita mengetahui target market yang sesuai untuk setiap kelompok.

[](https://www.dicoding.com/academies/184/tutorials/8412?from=8556)

Sebuah model pengklasteran akan membandingkan atribut setiap pengunjung lalu membuat sebuah klaster yang diisi oleh pengunjung yang memiliki kemiripan karakteristik/atribut yang tinggi.

[](https://www.dicoding.com/academies/184/tutorials/8412?from=8556)

Contoh di atas dikenal juga sebagai customer segmentation, salah satu kasus yang populer di industri, di mana bisnis mengelompokkan pelanggan agar bisa memberikan penawaran yang sesuai untuk setiap kelompok. Misal, kelompok pengunjung wanita dengan rentang usia 25 sampai 35 tahun tentu akan memiliki selera yang berbeda dengan pengunjung wanita pada rentang usia 40 tahun ke atas. Customer segmentation ini penting agar setiap target kelompok mendapatkan penawaran yang sesuai sehingga dapat memberikan kontribusi positif terhadap revenue toko.

Pada submodul selanjutnya kita akan belajar mengenai salah satu metode clustering yang sangat populer yaitu K-Means clustering.