|  |  |
| --- | --- |
| **M FARHAN MUZAKKI**  **1301154417**  **IF 39 11** | **MACHINE LEARNING**  **K MEANS CLUSTERING**  **LAPORAN** |

1. Analisis Masalah

K – Means Clustering adalah salah satu metode Clustering yang melakukan partisi data yang ada kedalam satu bentuk atau lebih kelompok. Centroid adalah titik yang menjadi pusat data. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya.

Pembelajaran ini termasuk dalam *unsupervised learning.* Masukan yang diterima adalah data atau objek dan *k* buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam *k* buah kelompok tersebut. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut.

1. Desain

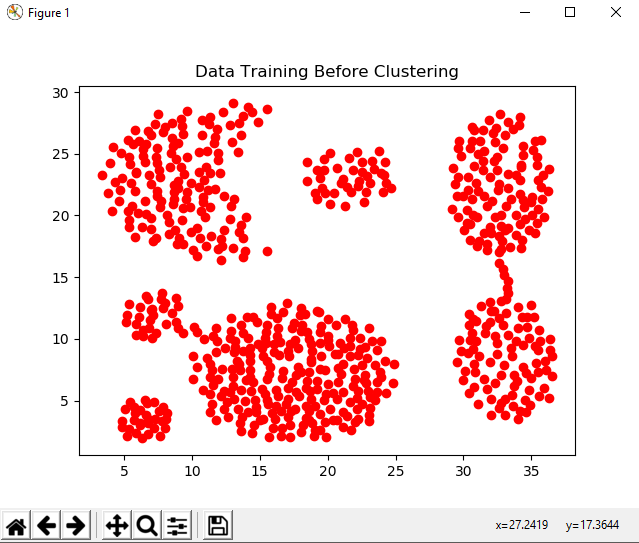
Berdasarkan tugas yang diberikan, terdapat data test dan data train yang harus di-Clustering. Pengerjaan tugas ini menggunakan Bahasa Pemrograman Python dan menggunakan compiler pycharm.

Tahapan dalam mengimplementasikan algoritma ini yaitu :

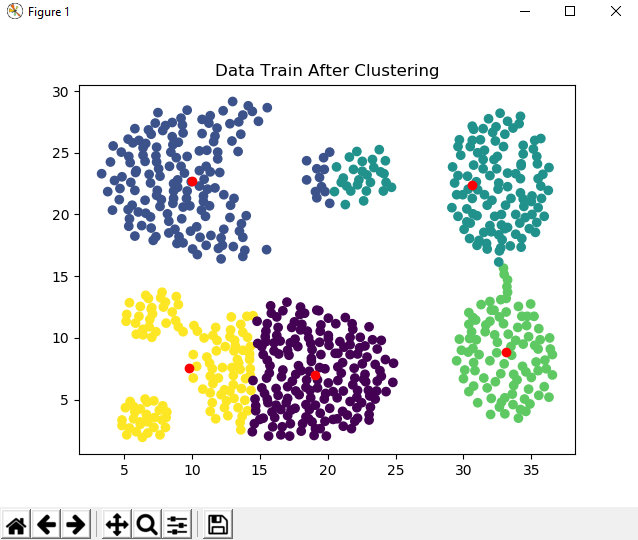
1. Menyiapkan data Train dan data Test
2. Set Nilai K Cluster
3. Set Nilai awal Centroid
4. Hitung Jarak antara Data centroid menggunakan rumus Euclideans Distance
5. Partisi Data berdasarkan nilai minimum
6. Melakukan Iterasi sealam partisi data masih bergerak
7. Bila Grup data tidak bergerak lagi, maka hentikan iterasi
8. Data telah berkelompok sesuai nilai centroid akhir

Hasil Scatter Plot :

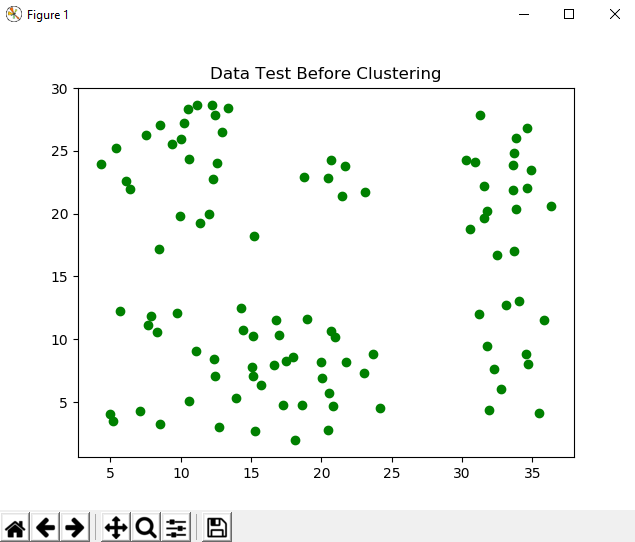
* Data Train Sebelum Clustering



* Data Train Sesudah Clustering



* Data Test Sebelum Clustering



* Data Test Sesudah Clustering



1. Evaluasi

Karena menentukan nilai Centroid dilakukan secara random, maka harus dipastikan bahwa nilai yang keluar tidaklah sama, karena akan menghasilkan NaN (not a number). Dalam menentukan jarak antar data Centroid dapat menggunakan cara selain menggunakan algoritma Euclidean Distance.