# Laporan klasifikasi *k-Means*

Dibuat untuk memenuhi tugas dua matakuliah pembelajaran mesin



Oleh:

Rama Aditya Maulana | 1301150034 | IF 39 06

**Fakultas Informatika**

**Universitas Telkom**

**Bandung**

**2018**

## Landasan Teori

Metode *k-means* merupakan algoritma klasterisasi yang paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena mudah diimplementasikan. Ide dari algoritma ini, yaitu meminimalkan *Sum Square Error* (SSE) antar objek-objek data dengan sejumlah *k centroid*.

* 1. Data yang akan diklasterisasi dipilih sejumlah *k* objek secara acak sebagai *centroid* awal.
  2. Setiap objek bukan *centroid* dimasukkan ke klaster terdekat berdasarkan urutan jarak tertentu.
  3. Setiap *centroid* diperbarui berdasarkan rata-rata objek yang ada didalam setiap klaster.
  4. Langkah kedua dan ketiga diiterasi hingga *centroid* stabil (konvergen)

## Rancangan Metode

Pada kasus ini, data terdiri dari dua buah berkas, yaitu

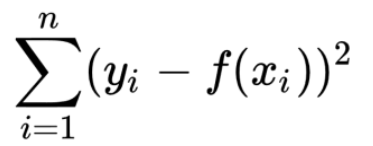
* 1. Data *Train*

*Data train* (data latih) yang disediakan berisi 688 data dalam *file* TrainingsetTugas2.txt

* 1. Data *Test*

*Data test* (data latih) yang disediakan berisi 100 data dalam *file* TestsetTugas2.txt

Pada program ini, nilai k (jumlah kelas) ditentukan menggunakan SSE agar optimal menggunakan metode *elbow*. Metode ini dihitung menggunakan rumus berikut.



SSE =

Metode *elbow* (siku) mengambil nilai titik balik dari penurunan SSE.

Setelah mendapatkan nilai SSE pada setiap jumlah cluster yang ditentukan.

Contoh grafik SSE seperti berikut.

Dari grafik diatas terlihat bahwa jumlah k optimum berada pada kelas tujuh dengan titik *centroid* awal adalah , sehingga kelas ketujuh dan titik *centroid* awal dipilih untuk menentukan klaster pada data *test*