

LAPORAN TUGAS 2
PEMBELAJARAN MESIN
K-MEANS CLUSTERING

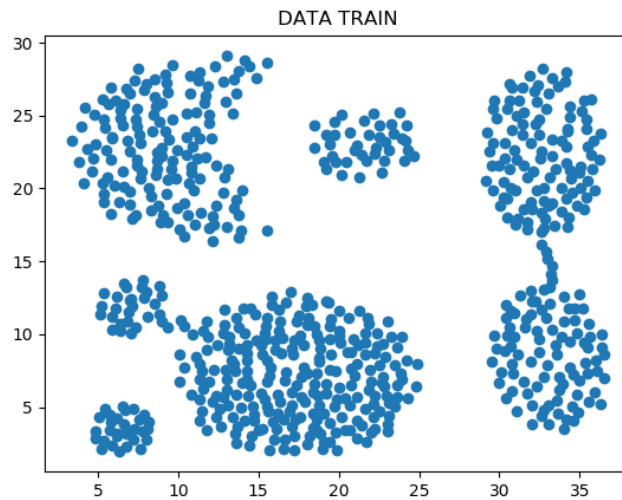


Disusun Oleh :
SEPTIAN DWI INDRADI
1301154164
IF-39-10

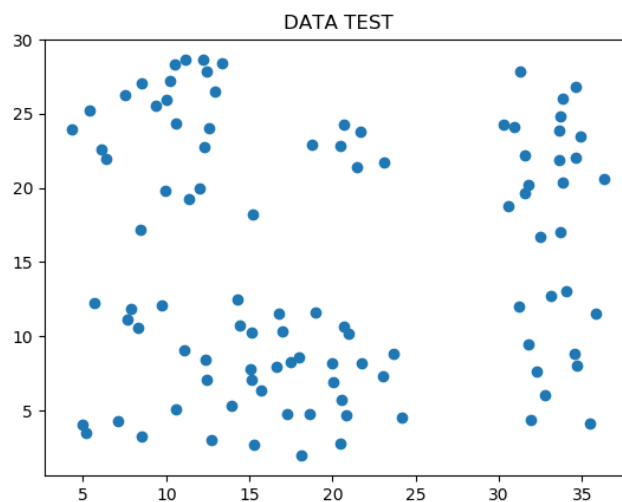
PROGRAM STUDI SARJANA INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS TELKOM
BANDUNG
2018

1. ANALISIS MASALAH

Diberikan 2 buah data yaitu Trainset yang terdiri dari 688 data dengan 2 atribut, dan Testset yang terdiri dari 100 data dengan 2 atribut, dimana kedua data tersebut tidak memiliki label. Untuk itu dibangun sebuah sistem untuk mengklasiterisasi (*clustering*) data tersebut.



Gambar 1. 1 Scatter plot Trainset

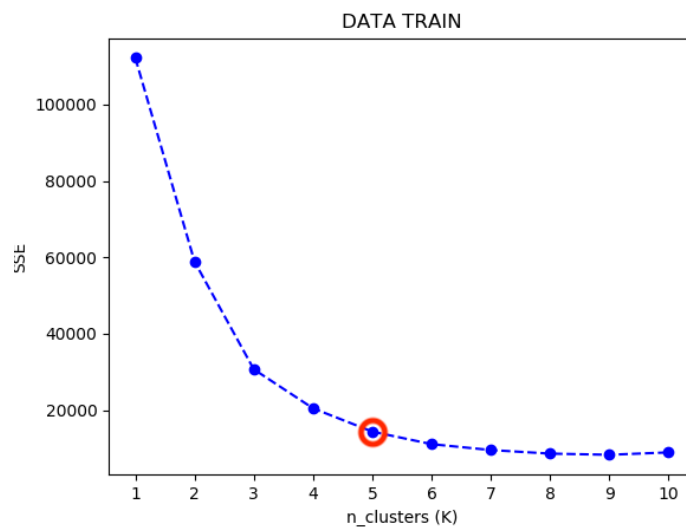


Gambar 1. 2 Scatter plot Testset

Gambar 1.1 dan gambar 1.2 merupakan *scatter plot* Testset dan Trainset yang belum memiliki cluster / label. Sistem *clustering* dibangun menggunakan algoritma K-Means yang merupakan algoritma *centroid-based clustering*. Parameter pada algoritma K-Means ini adalah menentukan K (jumlah cluster) yang tepat untuk data yang diberikan.

2. DESAIN SISTEM

Sistem *clustering* K-Means yang dibangun memiliki nilai $K=5$. Nilai K tersebut didapatkan dari observasi langsung dengan melihat persebaran data menggunakan *scatter plot* dan juga menggunakan *Elbow Method*, yaitu membandingkan SSE (*Sum Square Error*) dari *clustering* Trainset dengan nilai K yang diiterasi mulai dari $K=1$ hingga batas yang ditentukan. Pada kasus ini nilai K diiterasi dari 1 sampai 10 dan didapatkan hasil sebagai berikut :



Gambar 2. 1 Grafik SSE

$K = 1$; SSE = 112207.2074

$K = 2$; SSE = 58889.8070

$K = 3$; SSE = 30703.6119

$K = 4$; SSE = 24522.6070

$K = 5$; SSE = 14397.4971

$K = 6$; SSE = 11123.9569

$K = 7$; SSE = 9892.6731

$K = 8$; SSE = 8629.8078

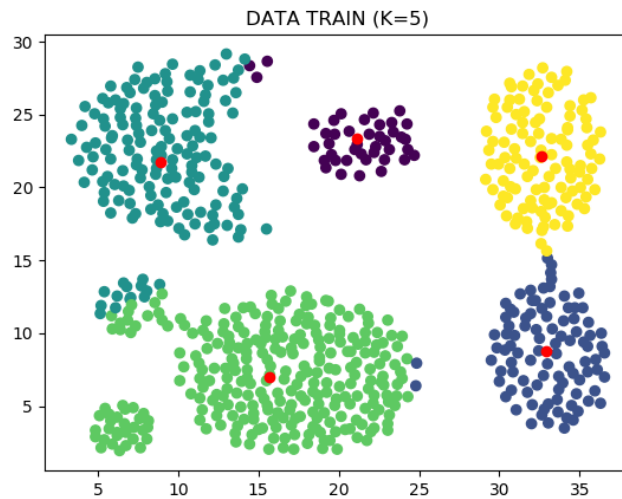
$K = 9$; SSE = 8396.2421

$K = 10$; SSE = 6974.2606

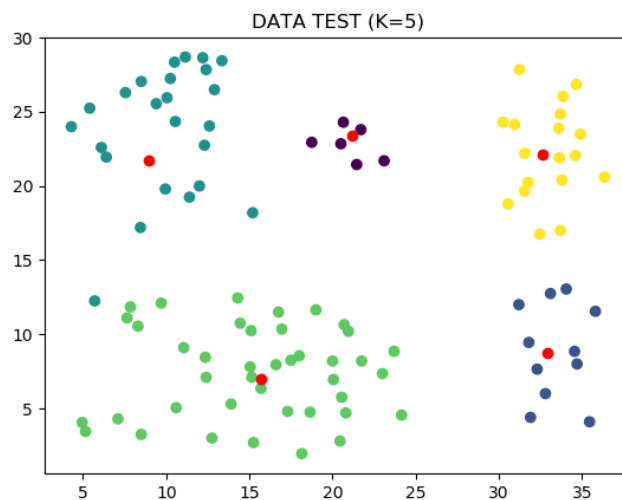
Dari data tersebut, K ditentukan dengan mengambil nilai K terkecil dimana SSE mulai turun secara stabil, yaitu pada $K=5$.

3. HASIL EKSPERIMEN

Berikut adalah hasil eksperimen *clustering* data menggunakan K-Means dengan $K=5$.



Gambar 3. 1 Hasil clustering trainset



Gambar 3. 2 Hasil clustering testset

Hasil *clustering* Trainset dan Testset berupa label dari setiap data kemudian disimpan dalam file “TrainsetLabel.txt” dan “TestsetLabel.txt”.