**K-MEANS CLUSTERING**

****

**Oleh :**

**Damaiyanto Hadomuan (1301150061)**

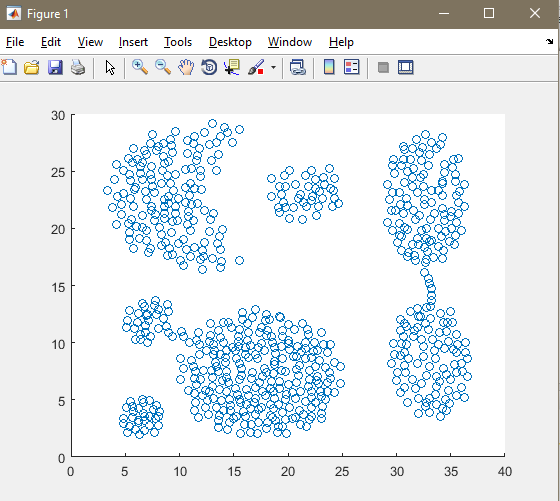
**IF 39-05**

Jl. Telekomunikasi No. 01, Terusan Buah Batu, Sukapura, Dayeuhkolot, Sukapura, Dayeuhkolot, Bandung, Jawa Barat 40257, Indonesia

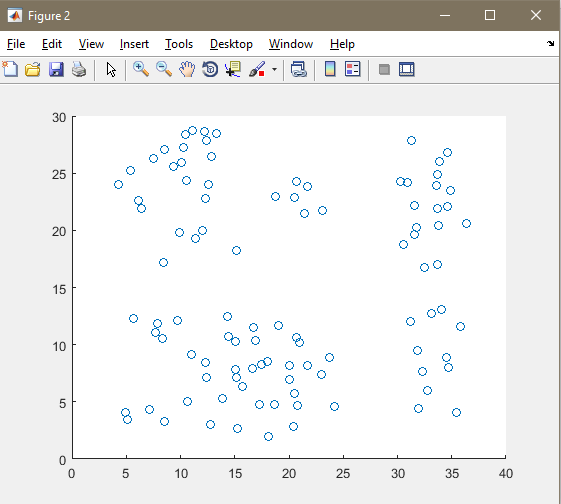
**2018**

**SOAL**

* PESEBARAN DATA TRAIN



* PESEBARAN DATA TEST

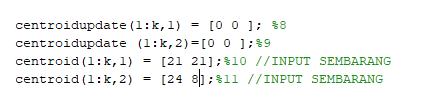


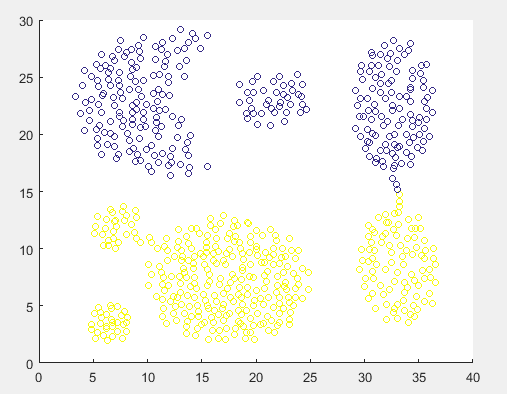
**ANALISIS**

Pada beberapa kali percobaan yang dilakukan dengan membagi beberapa cluster yang bermacam macam , yang menghasilkan pengelompokkan cluster sebagai berikut :

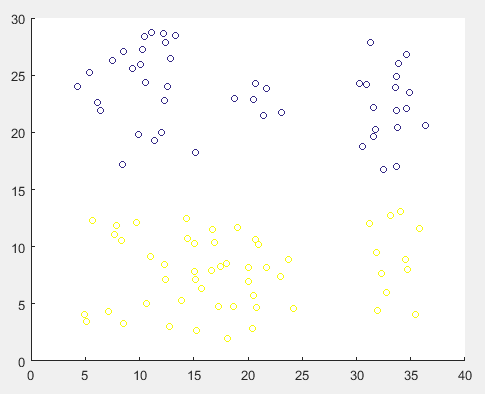
* Jika K=2

Centroid awal :



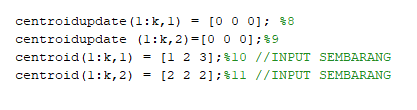


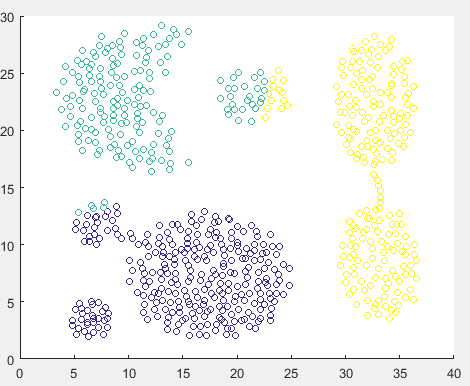
Pada pembagian 2 kluster , pembagian cenderung terbagi dengan baik , Namun memiliki kecenderungan SSE yang lebih besar



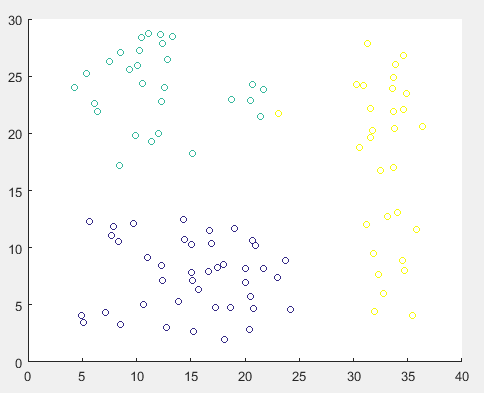
Centroid awal :

* JIKA K=3



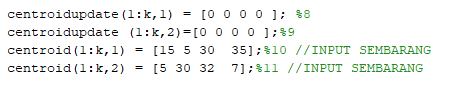


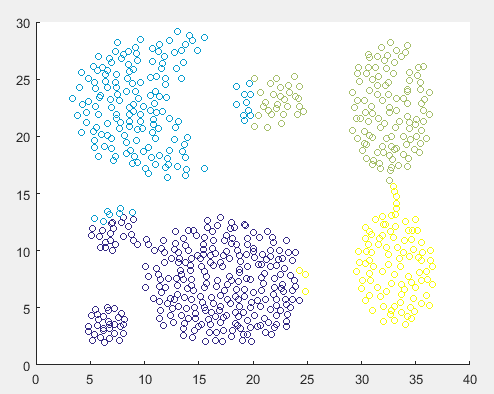
Pada pembagian 3 kluster , pembagian cenderung kurang baik terbagi dengan baik , Namun memiliki kecenderungan SSE yang lebih kecil dibandingkan pembagian 2 cluster

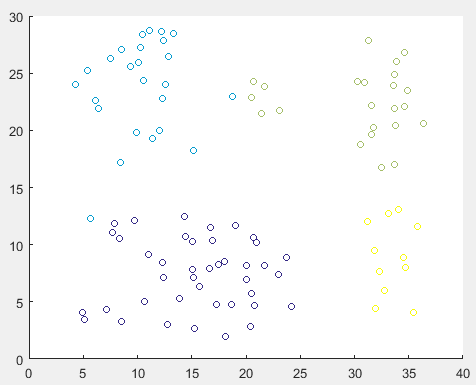


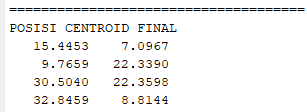
Jika K=4

Centroid awal :



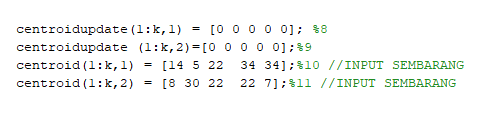


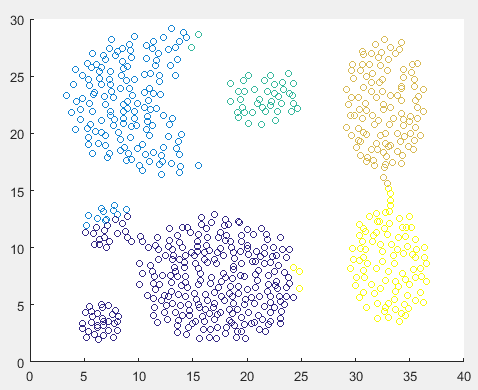


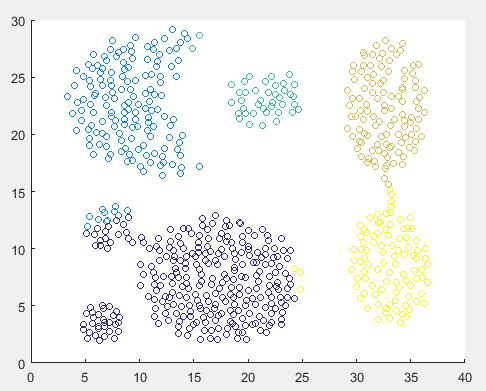
POSISI CENTROID FINAL

Jika K=5

Centroid awal :







Pada Pembagian 4 Cluster dan 5 Cluster , Keduanya kecenderungan memiliki SSE yang sama . Walaupun sudah dibagi 5 pada cluster terlihat pada gambar pesebaran data training , Noise yang berkurang pada setiap centroid jika dibandingkan dengan pembagian 4 cluster tidak terlalu signifikan , SEHINGGA lebih baik menggunakkan pembagian 4 cluster dibandingkan 5 cluster.

NOTE :

* Penentuan pembagian cluster yang baik dapat menggunakkan elbow , dimana dapat memilih yang memiliki rentang tidak terlalu curam dengan k sesudahnya.