

K-MEANS CLUSTERING



Oleh :

Damaiyanto Hedomuan (1301150061)

IF 39-05

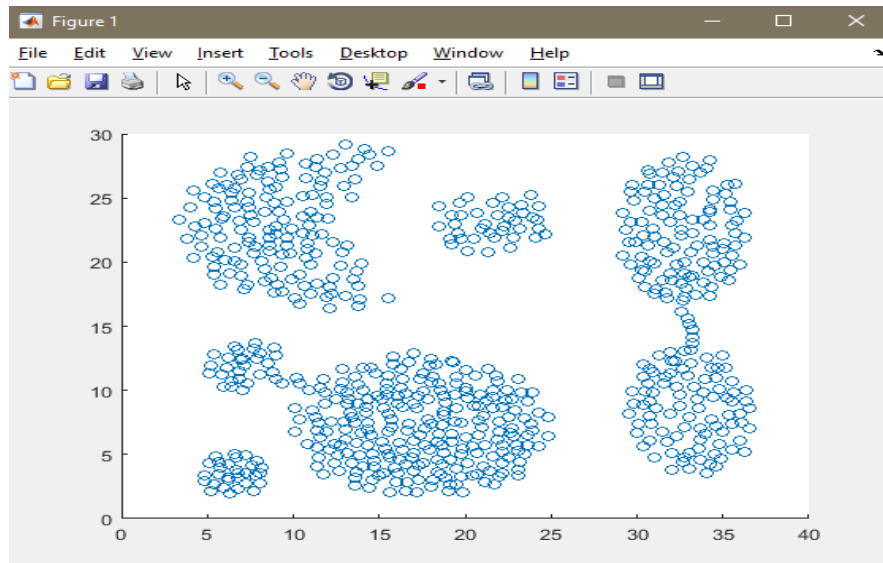
Jl. Telekomunikasi No. 01, Terusan Buah Batu, Sukapura, Dayeuhkolot, Sukapura, Dayeuhkolot,

Bandung, Jawa Barat 40257, Indonesia

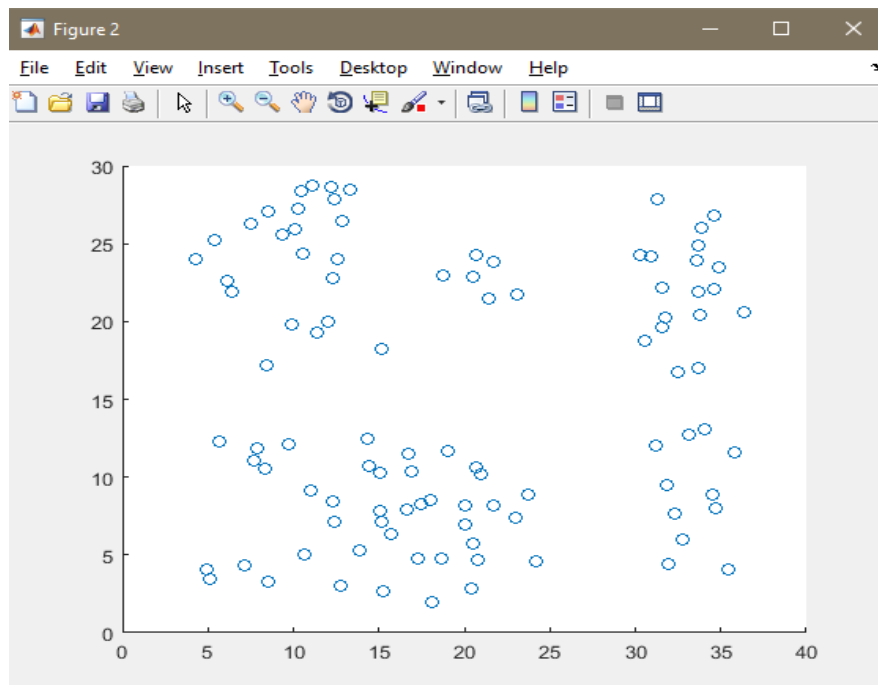
2018

SOAL

- PESEBARAN DATA TRAIN



- PESEBARAN DATA TEST



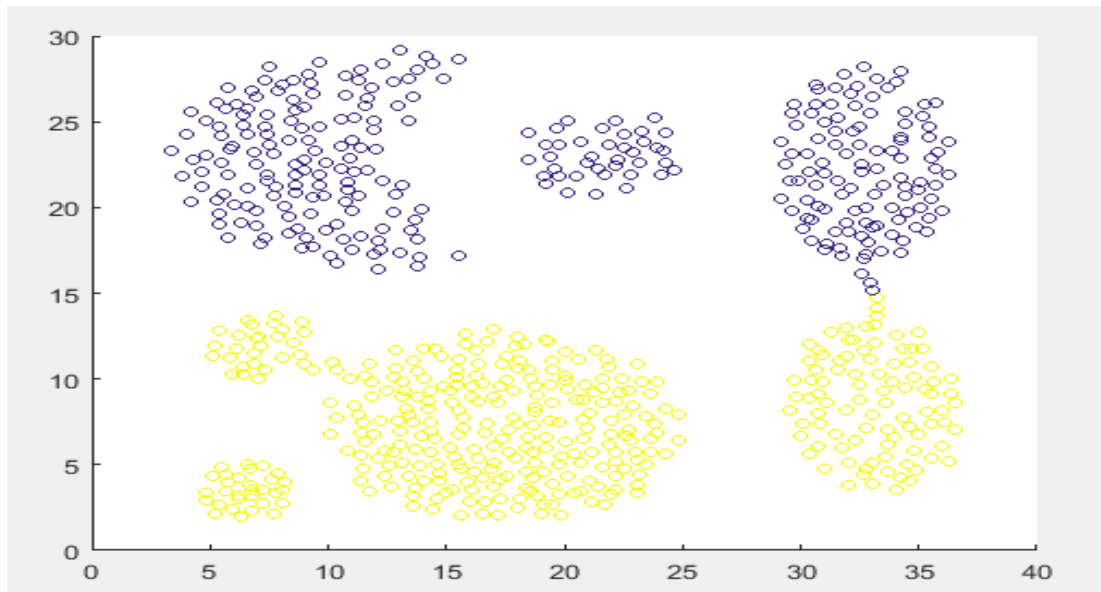
ANALISIS

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

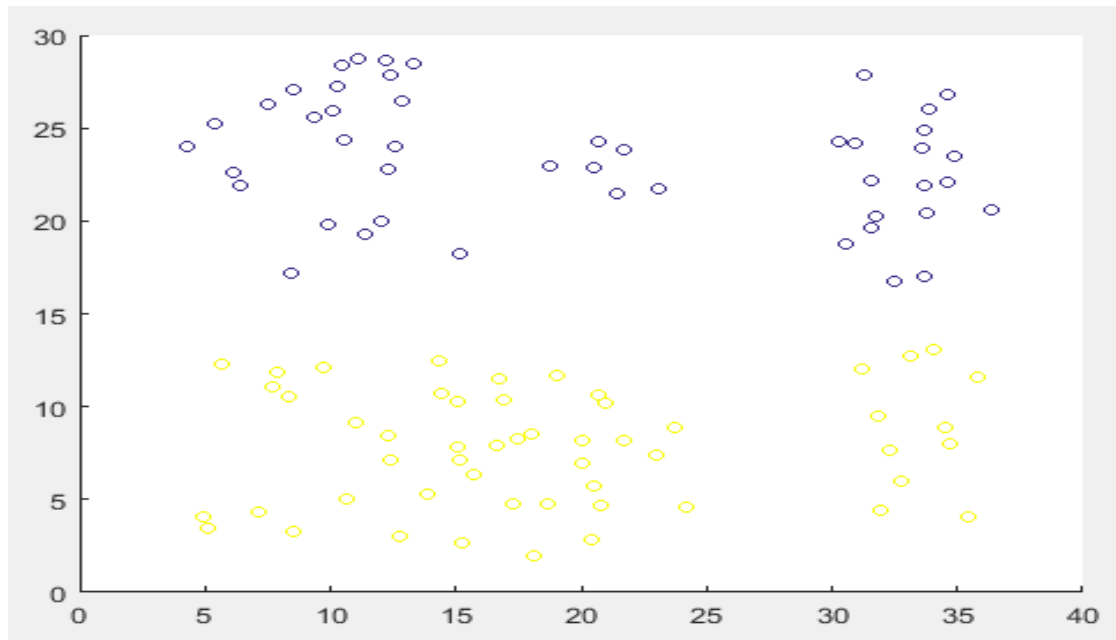
- Jika K=2

Centroid awal :

```
centroidupdate(1:k,1) = [0 0 ]; %8  
centroidupdate (1:k,2)=[0 0 ];%9  
centroid(1:k,1) = [21 21];%10 //INPUT SEMBARANG  
centroid(1:k,2) = [24 8];%11 //INPUT SEMBARANG
```



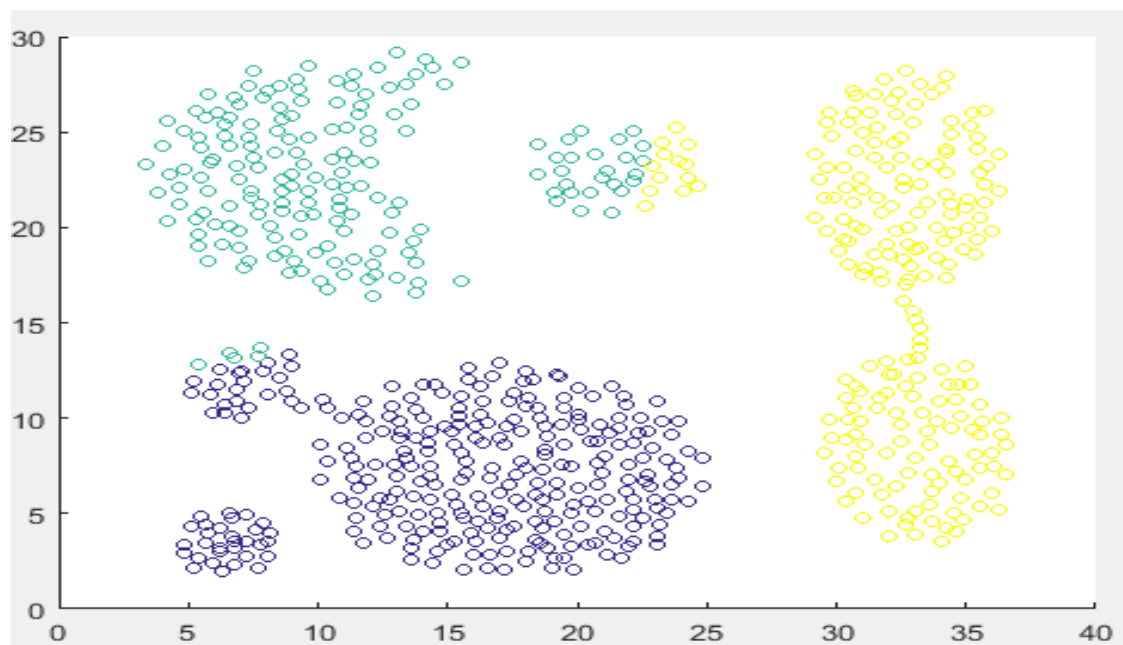
Pada pembagian 2 kluster , pembagian cenderung terbagi dengan baik , Namun memiliki kecenderungan SSE yang lebih besar



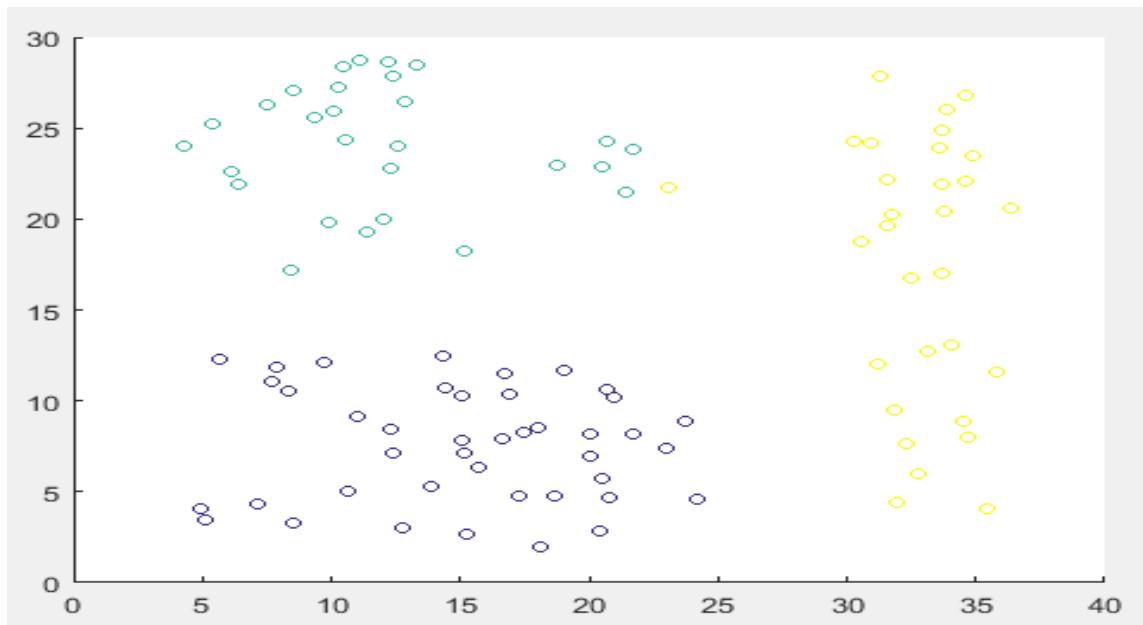
Centroid awal :

- JIKA K=3

```
centroidupdate(1:k,1) = [0 0 0]; %8
centroidupdate (1:k,2)=[0 0 0];%9
centroid(1:k,1) = [1 2 3];%10 //INPUT SEMBARANG
centroid(1:k,2) = [2 2 2];%11 //INPUT SEMBARANG
```



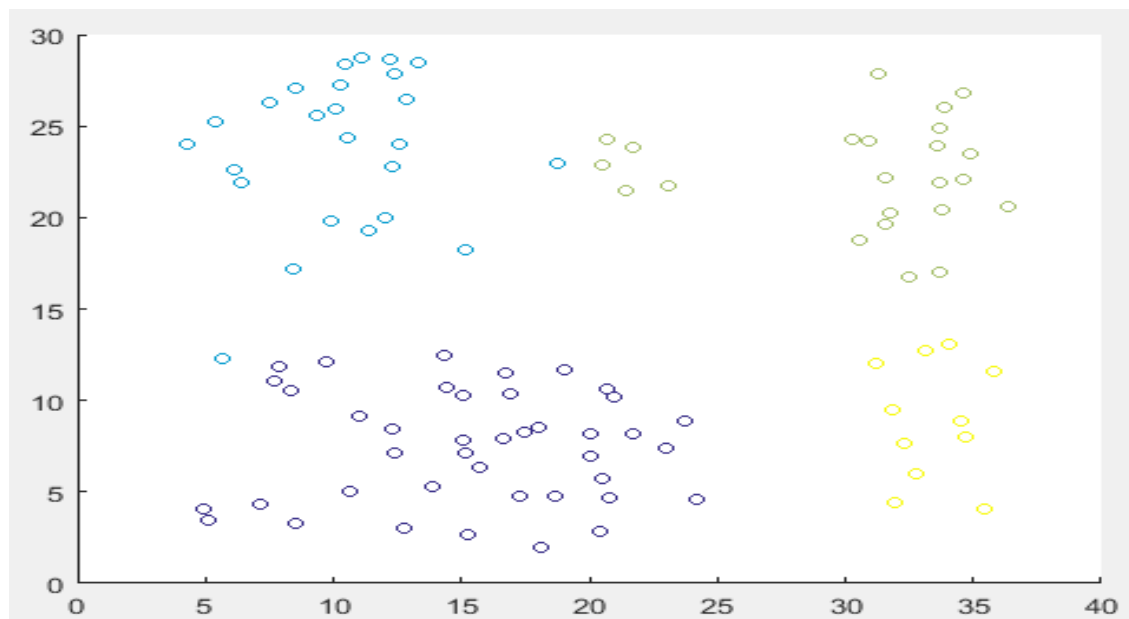
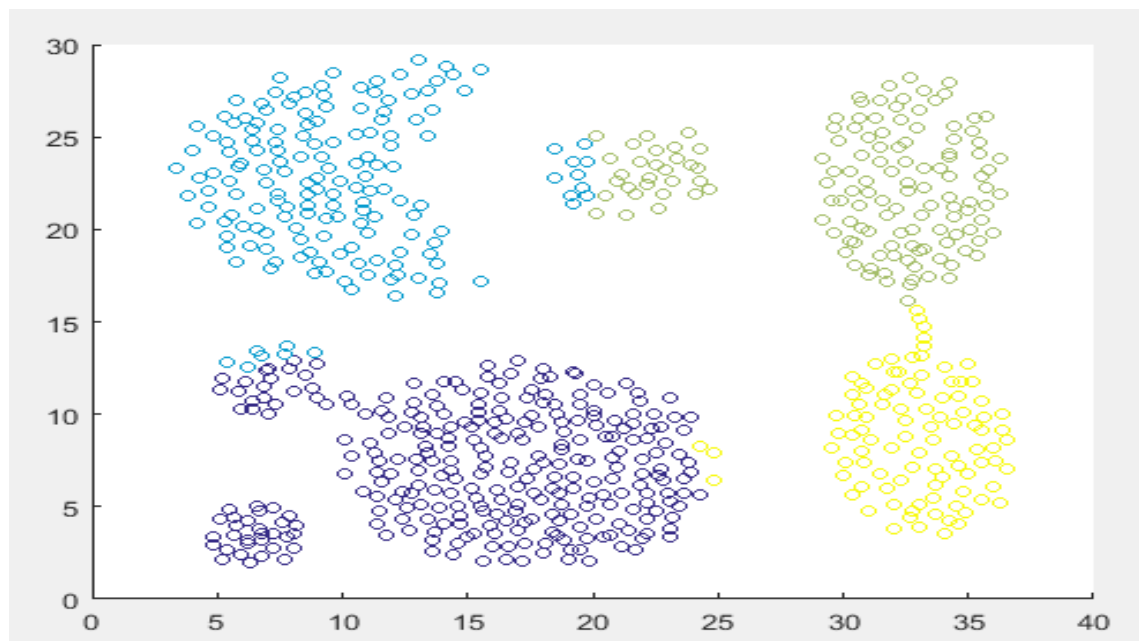
Pada pembagian 3 kluster , pembagian cenderung kurang baik terbagi dengan baik , Namun memiliki kecenderungan SSE yang lebih kecil dibandingkan pembagian 2 cluster



Jika K=4

Centroid awal :

```
centroidupdate(1:k,1) = [0 0 0 0 ]; %8  
centroidupdate (1:k,2)=[0 0 0 0 ];%9  
centroid(1:k,1) = [15 5 30 35];%10 //INPUT SEMBARANG  
centroid(1:k,2) = [5 30 32 7];%11 //INPUT SEMBARANG
```



POSISI CENTROID FINAL

=====

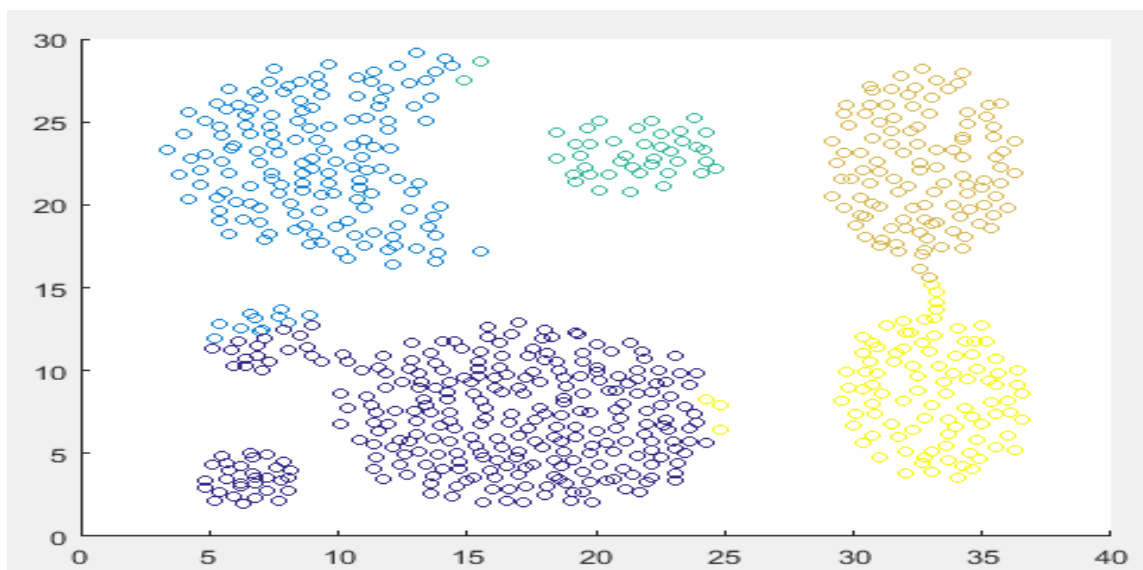
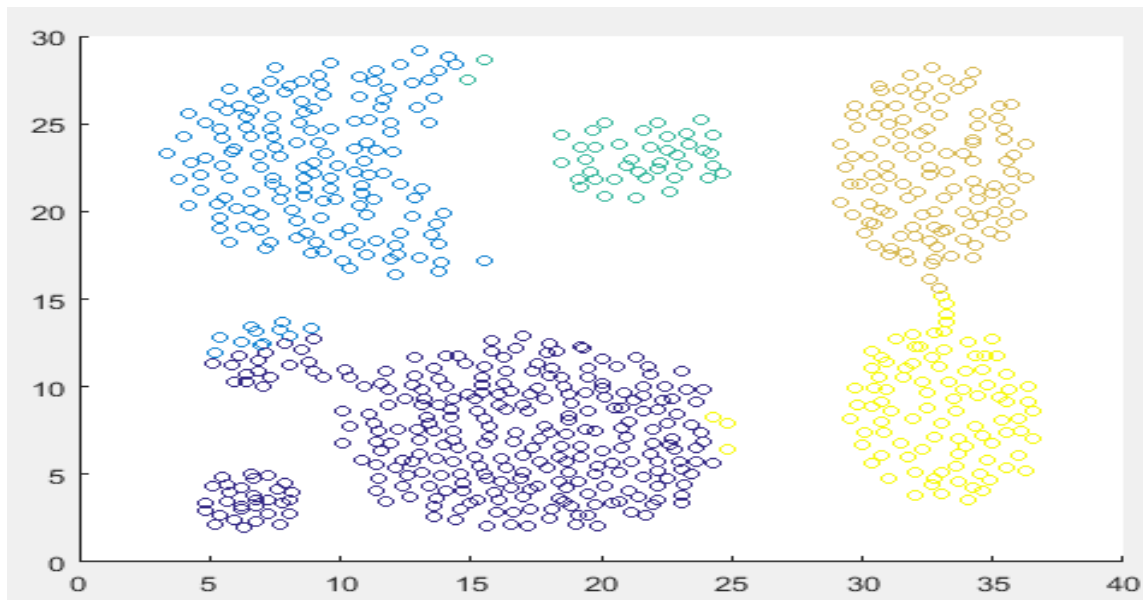
POSISI CENTROID FINAL

15.4453	7.0967
9.7659	22.3390
30.5040	22.3598
32.8459	8.8144

Jika K=5

Centroid awal :

```
centroidupdate(1:k,1) = [0 0 0 0 0]; %8  
centroidupdate (1:k,2)=[0 0 0 0 0];%9  
centroid(1:k,1) = [14 5 22 34 34];%10 //INPUT SEMBARANG  
centroid(1:k,2) = [8 30 22 22 7];%11 //INPUT SEMBARANG
```



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 menggunakan pembagian 4 cluster dibandingkan 5 cluster.

NOTE :

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