

MACHINE LEARNING

TUGAS 2 K-MEANS CLUSTERING

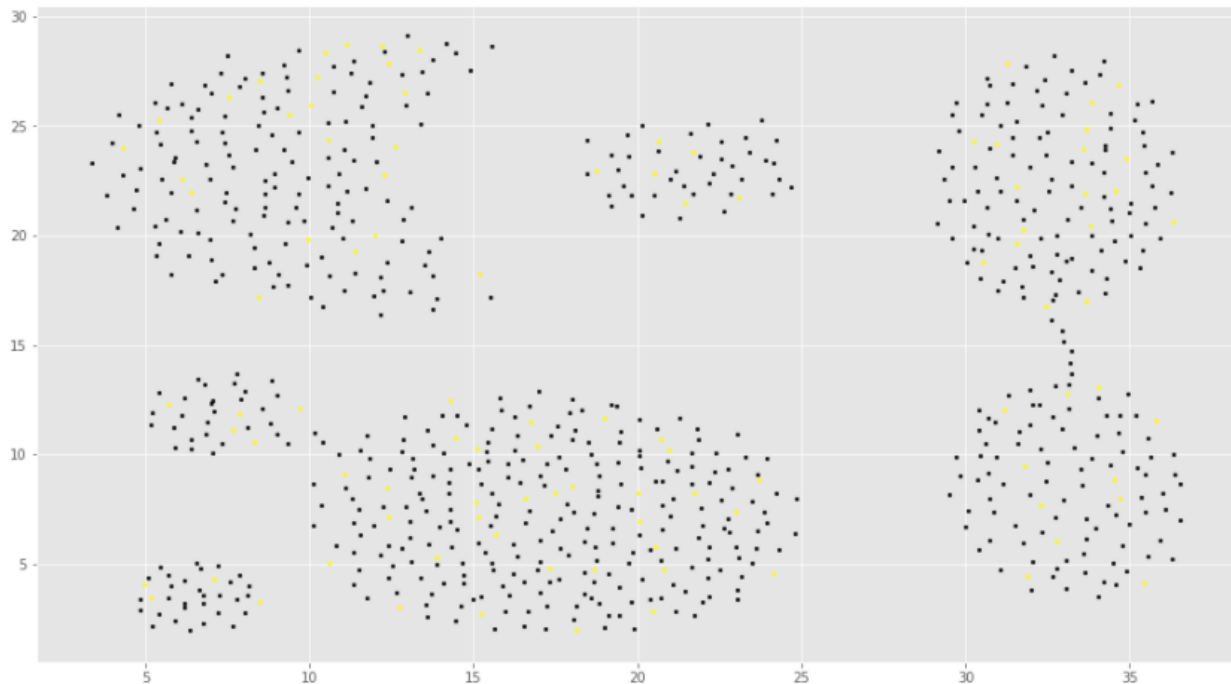
Nama : Astri Cahyaningtyas

NIM : 1301154548

Kelas : IF-39-02

ANALISIS MASALAH

K-Means adalah algoritma paling sederhana yang mengcluster data-data menjadi sebanyak K cluster. Tentukanlah cluster dari 100 data test yang dibantu dengan 600 data train.



Keterangan:

Hitam = train set

Kuning = test set

MACHINE LEARNING

TUGAS 2 K-MEANS CLUSTERING

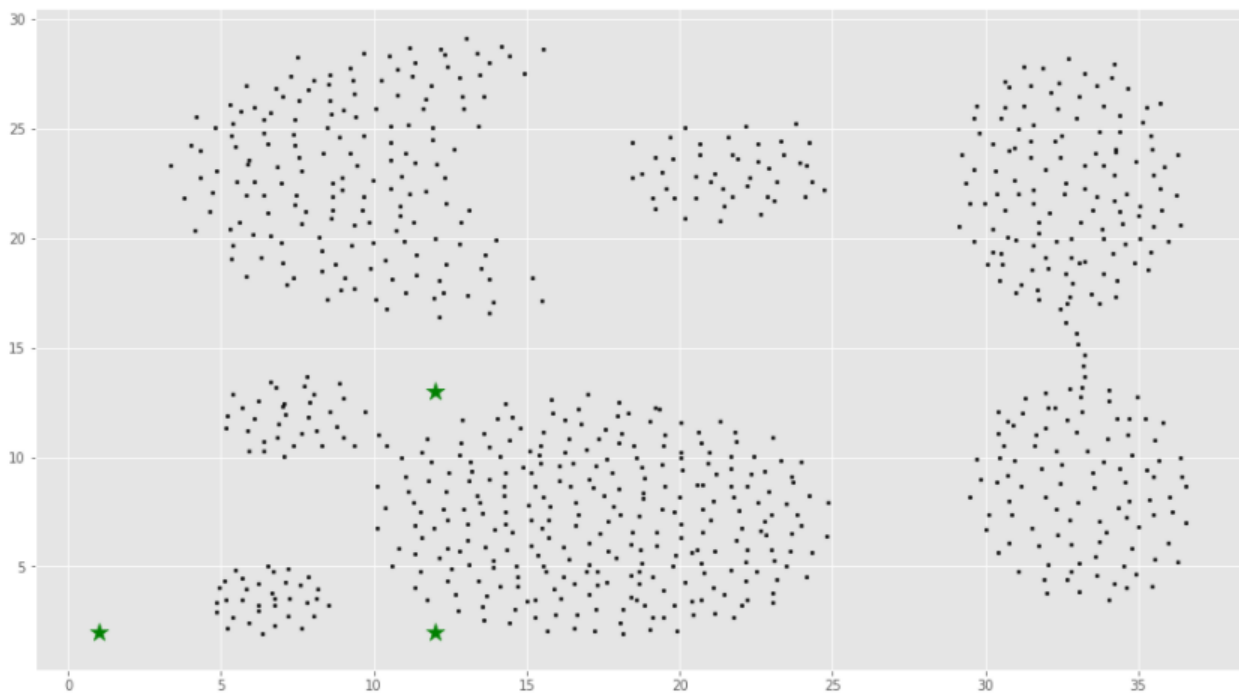
DESAIN

Langkah 1

Tentukan nilai centroid secara random. Asumsi C adalah himpunan nilai centroid. Nilai $k = 3$

```
# Banyaknya k cluster
k = 3
# X0 nilai centroid random
C_x0 = np.random.randint(0, np.max(X)-20, size=k)
# X1 nilai centroid random
C_x1 = np.random.randint(0, np.max(X)-20, size=k)
C = np.array(list(zip(C_x0, C_x1)), dtype=np.float32)
print(C)
```

```
[[ 1.  2.]
 [12. 13.]
 [12.  2.]]
```



Keterangan:

Bintang = nilai centroid awal/random

Langkah 2

Menghitung nilai euclidean (jarak antara titik dan setiap centroid).

MACHINE LEARNING

TUGAS 2 K-MEANS CLUSTERING

```
# Menghitung jarak Euclidean  
def dist(a, b, ax=1):  
    return np.linalg.norm(a - b, axis=ax)
```

Langkah 3

Menemukan nilai centroid baru dengan menghitung rata-rata dari semua titik di cluster tersebut.

```
# Menyimpan nilai centroid, selalu update  
C_old = np.zeros(C.shape)  
# Label(0, 1, 2)  
clusters = np.zeros(len(X))  
# Error func. Jarak antar centroid lama dna baru  
error = dist(C, C_old, None)  
# Loop sampai error 0  
while error != 0:  
    # Assigning each value to its closest cluster  
    for i in range(len(X)):  
        distances = dist(X[i], C)  
        cluster = np.argmin(distances)  
        clusters[i] = cluster  
    # Storing the old centroid values  
    C_old = deepcopy(C)  
    # Finding the new centroids by taking the average value  
    for i in range(k):  
        points = [X[j] for j in range(len(X)) if clusters[j] == i]  
        C[i] = np.mean(points, axis=0)  
    error = dist(C, C_old, None)
```

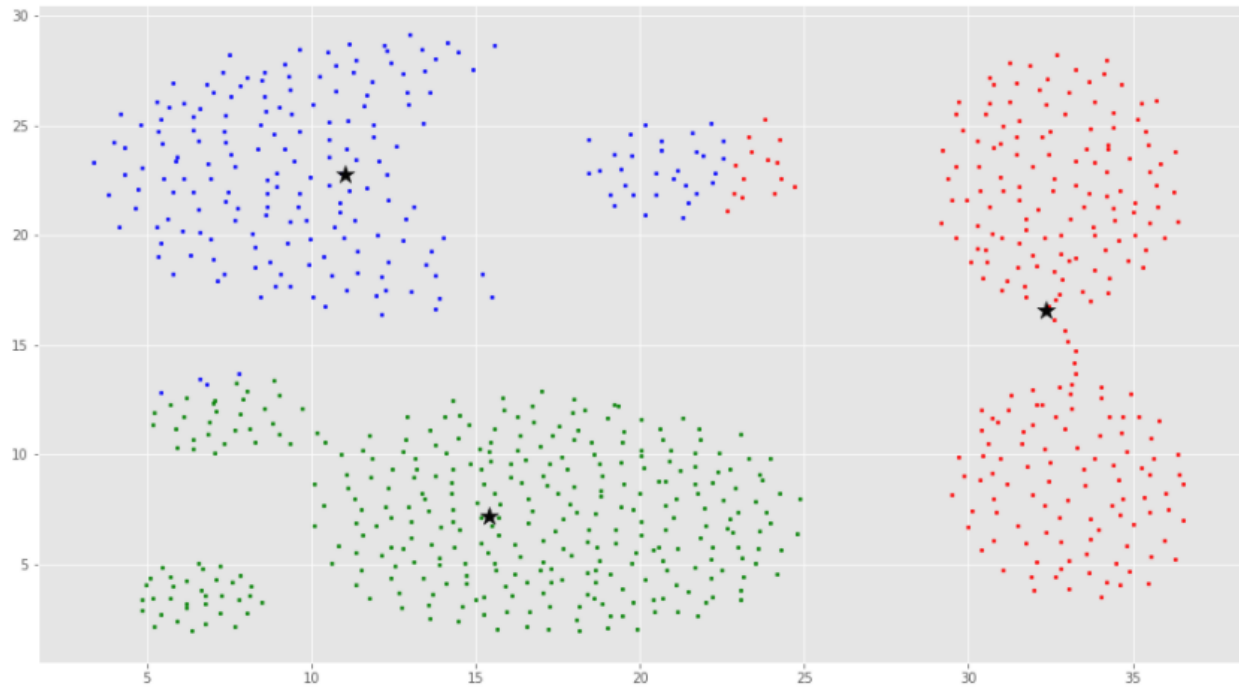
Langkah 4

Ulangi langkah 2 dan 3 sampai tidak ada perubahan pada nilai centroid

MACHINE LEARNING

TUGAS 2 K-MEANS CLUSTERING

EVALUASI HASIL EKSPERIMEN



Gambar diatas merupakan hasil eksperimen K-Means. Titik-titik yang berdekatan kemungkinan besar menjadi satu cluster. Berdasarkan hasil observasi cluster dari test set adalah sebagai berikut:

x0	x1	cluster
18.75	22.95	blue
21.45	21.45	blue
20.5	22.85	blue
20.65	24.3	blue
21.7	23.8	blue
23.1	21.7	green
13.35	28.45	blue
12.4	27.85	blue
12.2	28.65	blue
12.9	26.5	blue
11.15	28.7	blue
10.5	28.35	blue
10.25	27.25	blue
12.6	24.05	blue
10.05	25.95	blue
8.5	27.05	blue
7.55	26.3	blue

MACHINE LEARNING

TUGAS 2 K-MEANS CLUSTERING

9.4	25.55	blue
10.55	24.35	blue
5.4	25.25	blue
4.3	24	blue
6.1	22.6	blue
6.4	21.95	blue
8.45	17.2	blue
12.3	22.75	blue
9.95	19.8	blue
12	20	blue
11.4	19.25	blue
15.2	18.2	blue
31.9	4.4	green
32.8	6	green
35.45	4.1	green
32.3	7.65	green
31.8	9.45	green
34.7	8	green
34.55	8.85	green
31.2	12	green
33.1	12.75	green
35.8	11.55	green
34.05	13.05	green
11.05	9.1	red
14.3	12.45	red
14.45	10.75	red
12.35	8.45	red
12.4	7.1	red
15.1	10.25	red
16.75	11.5	red
15.05	7.8	red
10.6	5.05	red
13.9	5.3	red
12.75	3	red
15.15	7.1	red
15.25	2.7	red
15.7	6.35	red
17.3	4.8	red
18.15	1.95	red
20.45	2.8	red

MACHINE LEARNING

TUGAS 2 K-MEANS CLUSTERING

18.65	4.75	red
16.6	7.95	red
20.8	4.7	red
20.55	5.75	red
17.5	8.25	red
18	8.55	red
20.05	6.95	red
24.15	4.55	red
20	8.2	red
16.95	10.35	red
19	11.65	red
20.7	10.65	red
20.95	10.2	red
21.75	8.2	red
23	7.35	red
23.7	8.85	red
5.15	3.45	red
4.95	4.05	red
7.1	4.3	red
8.5	3.25	red
32.45	16.75	green
30.55	18.8	green
31.55	19.65	green
33.7	17	green
31.75	20.25	green
31.55	22.2	green
30.95	24.15	green
33.65	21.9	green
33.8	20.4	green
36.35	20.6	green
34.6	22.05	green
34.9	23.5	green
33.6	23.9	green
33.7	24.85	green
30.25	24.3	green
31.25	27.85	green
33.85	26.05	green
34.65	26.85	green
9.7	12.1	red
5.7	12.25	red

MACHINE LEARNING

TUGAS 2

K-MEANS CLUSTERING

7.85	11.85	red
7.65	11.1	red
8.3	10.55	red