



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
Bahasan : K-Means Clustering  
Halaman : 1/8

NIM	222410101064
Nama	Jetro Sulthan Fatih N
Kelas	A
Program Studi	Sistem Informasi
Asisten	1. Renata Sayidatul Arikha (212410101057) 2. Aprodhita Nanda Eka Wijaya (212410101071)

### LANGKAH KERJA

Lakukan clustering pada dataset seattle-weather. Langkah langkah

1. Normalisasikan dataset seattle-weather menggunakan normalisasi minmax

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt

df = pd.read_csv("seattle-weather.csv")

scaler = MinMaxScaler()
scaler.fit(df[['precipitation']])
df['precipitation'] = scaler.transform(df[['precipitation']])

scaler.fit(df[['temp_max']])
df['temp_max'] = scaler.transform(df[['temp_max']])

scaler.fit(df[['temp_min']])
df['temp_min'] = scaler.transform(df[['temp_min']])

scaler.fit(df[['wind']])
df['wind'] = scaler.transform(df[['wind']])

df.head()
```



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
Bahasan : K-Means Clustering  
Halaman : 2/8

	date	precipitation	temp_max	temp_min	wind	weather
0	2012-01-01	0.000000	0.387097	0.476378	0.472527	drizzle
1	2012-01-02	0.194991	0.327957	0.389764	0.450549	rain
2	2012-01-03	0.014311	0.357527	0.562992	0.208791	rain
3	2012-01-04	0.363148	0.370968	0.500000	0.472527	rain
4	2012-01-05	0.023256	0.282258	0.389764	0.626374	rain

### 2. Tampilkan scatter plot sebelum clustering

```
plt.scatter(df['precipitation'],df['temp_max'])  
plt.xlabel('Precipitation')  
plt.ylabel('Temp Max')
```

```
plt.scatter(df['temp_min'],df['wind'])  
plt.xlabel('Temp Min')  
plt.ylabel('Wind')
```

```
plt.scatter(df['precipitation'],df['temp_min'])  
plt.xlabel('Precipitation')  
plt.ylabel('Temp Min')
```

```
plt.scatter(df['temp_max'],df['wind'])  
plt.xlabel('Temp Max')  
plt.ylabel('Wind')
```



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
Bahasan : K-Means Clustering  
Halaman : 3/8

3. Clusterlah dataset seattle-weather dengan nilai K 2, 3, 4

```
km = KMeans(n_clusters=2)
# km = KMeans(n_clusters=3)
# km = KMeans(n_clusters=4)

y_predicted = km.fit_predict(df[['precipitation', 'temp_max']])
# y_predicted = km.fit_predict(df[['temp_min', 'wind']])
# y_predicted = km.fit_predict(df[['precipitation', 'temp_min']])
# y_predicted = km.fit_predict(df[['temp_max', 'wind']])
y_predicted

df['cluster'] = y_predicted
df.head()
km.cluster_centers_

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
# df3 = df[df.cluster==2]
# df4 = df[df.cluster==3]
```

4. Tentukan K terbaik jika dilakukan dengan analisis Elbow Method

```
dframe = df[['precipitation', 'temp_max']]
# dframe = df[['temp_min', 'wind']]
# dframe = df[['precipitation', 'temp_min']]
# dframe = df[['temp_max', 'wind']]
```



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
Bahasan : K-Means Clustering  
Halaman : 4/8

```
X = range(2,11)

inertia = []

for i in X:

    clustering = KMeans(n_clusters=i)

    clustering.fit(dframe)

    inertia.append(clustering.inertia_)

plt.plot(X, inertia, 'o-')
plt.xlabel('Nilai Range X :')
plt.ylabel('Inertia')
plt.title('The Elbow Method menggunakan Inertia')
plt.show()
```

### 5. Tampilkan scatter plot setelah clustering

```
plt.scatter(df1['precipitation'],df1['temp_max'],color='green')
plt.scatter(df2['precipitation'],df2['temp_max'],color='red')
# plt.scatter(df3['precipitation'],df3['temp_max'],color='blue')
# plt.scatter(df4['precipitation'],df4['temp_max'],color='black')
plt.scatter(km.cluster_centers_[0], km.cluster_centers_[1], color='purple',
marker='*', label='centroid')
plt.xlabel('Precipitation')
plt.ylabel('Temp Max')
plt.legend()

plt.scatter(df1['temp_min'],df1['wind'],color='green')
plt.scatter(df2['temp_min'],df2['wind'],color='red')
# plt.scatter(df3['temp_min'],df3['wind'],color='blue')
```



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
Bahasan : K-Means Clustering  
Halaman : 5/8

```
# plt.scatter(df4['temp_min'],df4['wind'],color='black')

plt.scatter(km.cluster_centers_[:,0], km.cluster_centers_[:,1], color='purple',
marker='*', label='centroid')

plt.xlabel('Temp Min')
plt.ylabel('Wind')
plt.legend()

plt.scatter(df1['precipitation'],df1['temp_min'],color='green')
plt.scatter(df2['precipitation'],df2['temp_min'],color='red')
# plt.scatter(df3['precipitation'],df3['temp_min'],color='blue')
# plt.scatter(df4['precipitation'],df4['temp_min'],color='black')

plt.scatter(km.cluster_centers_[:,0], km.cluster_centers_[:,1], color='purple',
marker='*', label='centroid')

plt.xlabel('Precipitation')
plt.ylabel('Temp Min')
plt.legend()

plt.scatter(df1['temp_max'],df1['wind'],color='green')
plt.scatter(df2['temp_max'],df2['wind'],color='red')
# plt.scatter(df3['temp_max'],df3['wind'],color='blue')
# plt.scatter(df4['temp_max'],df4['wind'],color='black')

plt.scatter(km.cluster_centers_[:,0], km.cluster_centers_[:,1], color='purple',
marker='*', label='centroid')

plt.xlabel('Temp Max')
plt.ylabel('Wind')
plt.legend()
```

6. Analisislah perbandingan scatter plot sebelum clustering dan setelah clustering



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
K-Means Clustering  
Bahasan :  
Halaman : 6/8

Sebelum melakukan clustering, scatter plot hanya menampilkan satu kelompok data dengan warna biru. Namun, setelah melakukan clustering dan menemukan pusat-pusatnya, scatter plot dari data cuaca Seattle menjadi terbagi-bagi sesuai dengan jumlah klaster (K) yang ditentukan. Misalnya, jika  $K=2$ , maka scatter plot akan terbagi menjadi dua kelompok dengan dua warna yang berbeda. Begitu juga jika  $K=3$  atau  $K=4$ , plot akan terbagi menjadi tiga atau empat kelompok, masing-masing dengan warna yang berbeda.

Dengan adanya proses clustering ini, kita dapat lebih memahami karakteristik dari setiap kelompok data. Setiap klaster menjadi lebih jelas dan memiliki ciri khasnya sendiri, dan tidak ada lagi data yang termasuk dalam kategori ambigu atau memiliki ciri-ciri yang ambigu antara satu klaster dengan klaster lainnya.

7. Tentukan manakah k terbaik jika dilakukan analisis secara visual

Tugas no 1 – 6 dilakukan terhadap kombinasi atribut

- precipitation - temp\_max
- temp\_min - wind
- precipitation ~ temp\_min
- temp\_max ~ wind

Secara visual, analisis menunjukkan bahwa K terbaik dalam kasus ini adalah  $K = 4$ . Hal ini didukung oleh Elbow Method yang menunjukkan penurunan inersia yang signifikan hanya hingga  $K = 4$ . Selain itu, scatter plot pada  $K = 4$  menunjukkan struktur plot yang lebih jelas dan mudah dipahami. Dengan menggunakan  $K = 4$ , kita dapat dengan lebih tepat mengidentifikasi karakteristik dari setiap klaster, tanpa adanya keambiguan yang signifikan atau spesifikasi yang berlebihan.

### HASIL DAN ANALISIS DATA

Kami melakukan normalisasi Min-Max terhadap dataset cuaca Seattle yang diperoleh dari praktikum menggunakan file seattle-weather.csv. Normalisasi ini bertujuan untuk mengubah nilai-nilai angka dalam dataset menjadi nilai float dengan rentang dari 0 hingga 1, sehingga memudahkan proses clustering nantinya.

Setelah normalisasi, kami menampilkan scatter plot untuk setiap kombinasi antara atribut precipitation, temp\_min, temp\_max, dan wind sebelum melakukan clustering. Hasilnya adalah plot dengan titik-titik yang merepresentasikan setiap data dalam dataset.



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
Bahasan : K-Means Clustering  
Halaman : 7/8

Selanjutnya, kami melakukan proses clustering terhadap dataset dengan nilai K sebesar 2, 3, dan 4. Kami mencari nilai K yang paling efektif dengan menggunakan Elbow Method, yang menghitung inersia dari  $K = 1$  hingga  $K = n$ . Ditemukan bahwa puncak penurunan inersia drastis berada di  $K = 4$ .

Kemudian, kami menampilkan scatter plot untuk setiap nilai K yang berbeda, yaitu  $K = 2$ ,  $K = 3$ , dan  $K = 4$ . Kami menganalisis perbedaan antara scatter plot data sebelum clustering dan setelah clustering dilakukan untuk setiap nilai K.

Berdasarkan analisis visual terhadap output dari scatter plot, kami menemukan bahwa nilai K terbaik adalah  $K = 4$ . Ini disimpulkan dari penampilan scatter plot yang menunjukkan struktur plot yang lebih jelas dan mudah dipahami setelah dilakukan clustering dengan  $K = 4$ .

### KESIMPULAN

Kami mempelajari secara praktis cara kerja algoritma K-Means Clustering dan dampaknya terhadap visualisasi data. Kami juga memahami bagaimana menambahkan dan mengatur nilai K untuk menciptakan kluster sesuai dengan centroid yang ditentukan secara otomatis. Selain itu, kami memahami fungsi dari Elbow Method, yaitu untuk menemukan titik di mana penurunan inersia drastis terakhir terjadi, yang membantu kita menentukan nilai K terbaik untuk clustering data.

Link Google Colab

 LKM5\_1064.ipynb

Link Youtube (Unlisted)

[https://youtu.be/M0\\_1\\_xGgzzk](https://youtu.be/M0_1_xGgzzk)

Jember, .....2024

Mengetahui,  
Dosen Datamining

Asisten,

Fajrin Nurman Arifin, S.T., M.Eng

(Nama Jelas)



## Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining  
Bahasan : K-Means Clustering  
Halaman : 8/8

NIP. 198511282015041002

NIM.