



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 1/15

NIM	222410101095
Nama	Muhammad Afif Rohman Muzaky
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()
```

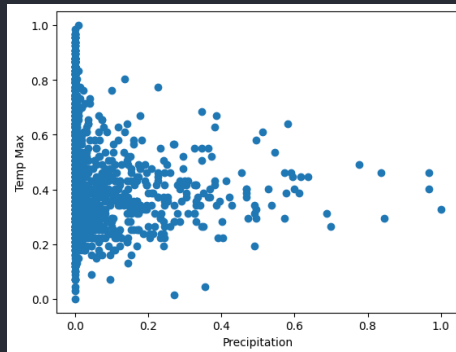
2. Tampilkan scatter plot sebelum clustering



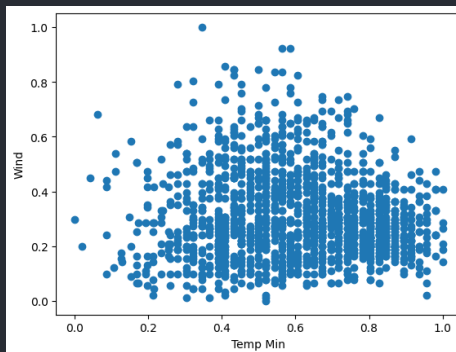
Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 2/15

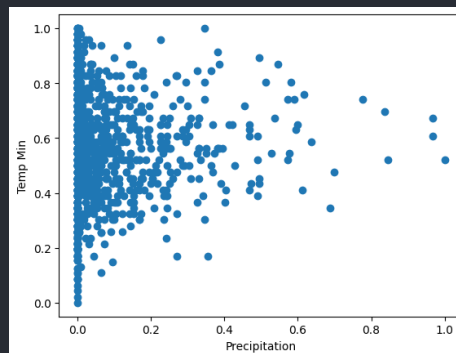
```
plt.scatter(df['precipitation'],df['temp_max'])  
plt.xlabel('Precipitation')  
plt.ylabel('Temp Max')
```



```
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')
```

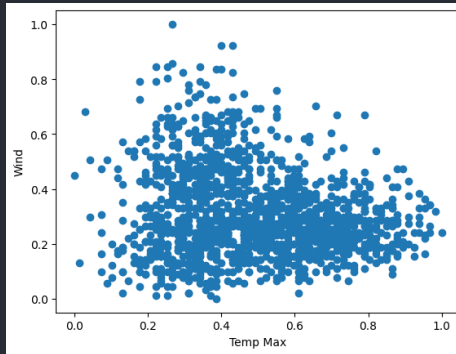




Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 3/15

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



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

```
kmean = KMeans(n_clusters=2)  
# kmean = KMeans(n_clusters=3)  
# kmean = KMeans(n_clusters=4)  
  
y_predicted = kmean.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()  
kmean.cluster_centers_  
  
df1 = df[df.cluster==0]  
# df2 = df[df.cluster==1]  
# df3 = df[df.cluster==2]  
# df4 = df[df.cluster==3]
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 4/15

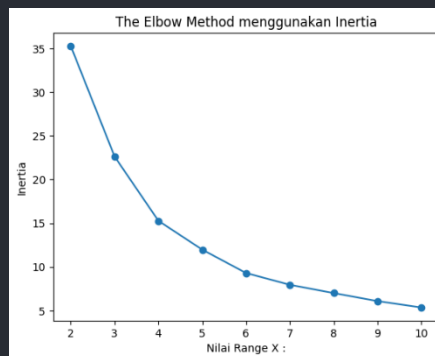
4. Tentukan K terbaik jika dilakukan dengan analisis Elbow Method

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

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

#Cluster 2

```
kmean = KMeans(n_clusters=2)
y_predicted = kmean.fit_predict(df[['precipitation', 'temp_max']])
y_predicted
```

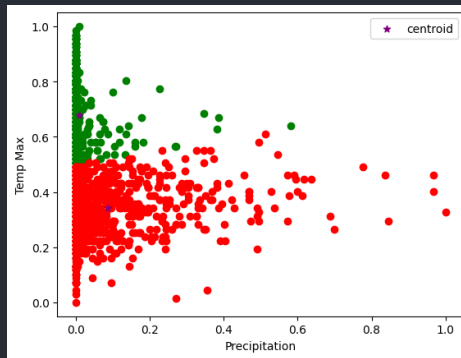


Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 5/15

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

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
plt.scatter(df1['precipitation'],df1['temp_max'],color='green')
plt.scatter(df2['precipitation'],df2['temp_max'],color='red')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Precipitation')
plt.ylabel('Temp Max')
plt.legend()
```



```
kmean = KMeans(n_clusters=2)
y_predicted = kmean.fit_predict(df[['temp_min', 'wind']])
y_predicted

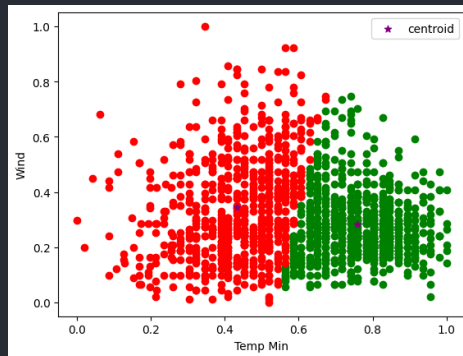
df['cluster'] = y_predicted
df.head()
kmean.cluster_centers_

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
plt.scatter(df1['temp_min'],df1['wind'],color='green')
plt.scatter(df2['temp_min'],df2['wind'],color='red')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Temp Min')
plt.ylabel('Wind')
plt.legend()
```



Lembar Kerja Mahasiswa

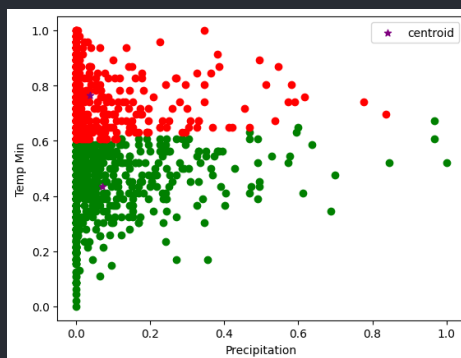
Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 6/15



```
kmean = KMeans(n_clusters=2)
y_predicted = kmean.fit_predict(df[['precipitation', 'temp_min']])
y_predicted

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

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
plt.scatter(df1['precipitation'], df1['temp_min'], color='green')
plt.scatter(df2['precipitation'], df2['temp_min'], color='red')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
            marker='*', label='centroid')
plt.xlabel('Precipitation')
plt.ylabel('Temp Min')
plt.legend()
```



```
kmean = KMeans(n_clusters=2)
y_predicted = kmean.fit_predict(df[['temp_max', 'wind']])
y_predicted

df['cluster'] = y_predicted
```

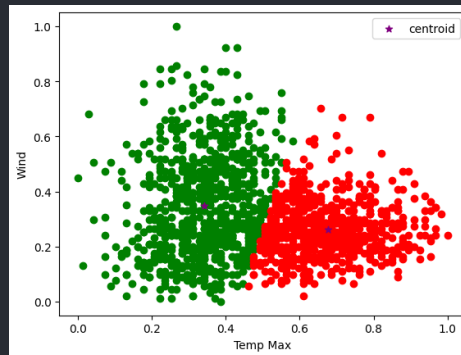


Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 7/15

```
df.head()
kmean.cluster_centers_

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
plt.scatter(df1['temp_max'],df1['wind'],color='green')
plt.scatter(df2['temp_max'],df2['wind'],color='red')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Temp Max')
plt.ylabel('Wind')
plt.legend()
```



#Cluster 3

```
kmean = KMeans(n_clusters=3)
y_predicted = kmean.fit_predict(df[['precipitation', 'temp_max']])
y_predicted

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

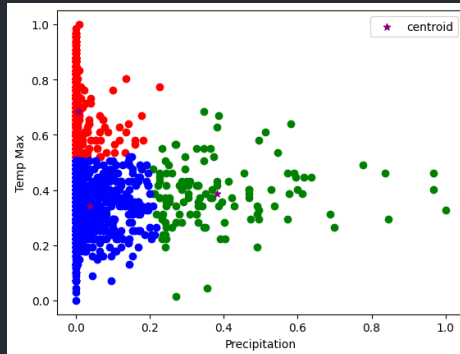
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1['precipitation'],df1['temp_max'],color='green')
plt.scatter(df2['precipitation'],df2['temp_max'],color='red')
plt.scatter(df3['precipitation'],df3['temp_max'],color='blue')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Precipitation')
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 8/15

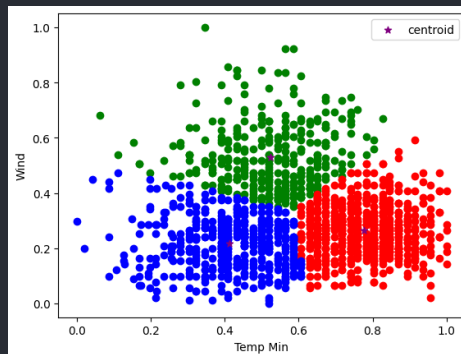
```
plt.ylabel('Temp Max')  
plt.legend()
```



```
kmean = KMeans(n_clusters=3)  
y_predicted = kmean.fit_predict(df[['temp_min', 'wind']])  
y_predicted
```

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

```
df1 = df[df.cluster==0]  
df2 = df[df.cluster==1]  
df3 = df[df.cluster==2]  
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')  
plt.scatter(kmean.cluster_centers_[0,0], kmean.cluster_centers_[0,1], color='purple',  
            marker='*', label='centroid')  
plt.xlabel('Temp Min')  
plt.ylabel('Wind')  
plt.legend()
```



```
kmean = KMeans(n_clusters=3)  
y_predicted = kmean.fit_predict(df[['precipitation', 'temp_min']])  
y_predicted
```

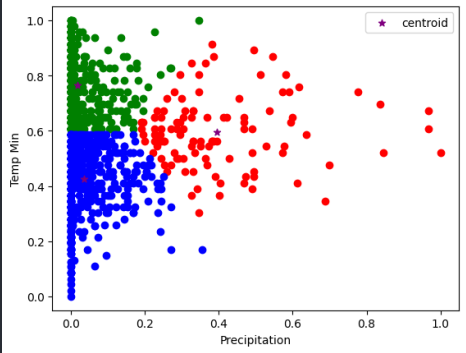



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 9/15

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

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
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(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Precipitation')
plt.ylabel('Temp Min')
plt.legend()
```



```
kmean = KMeans(n_clusters=3)
y_predicted = kmean.fit_predict(df[['temp_max', 'wind']])
y_predicted

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

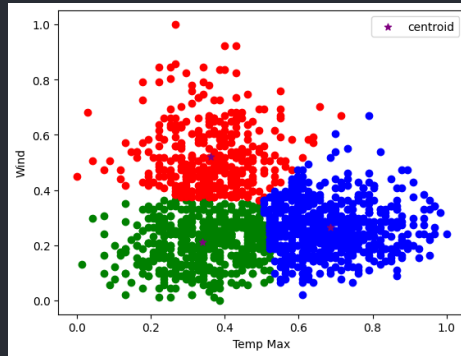
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1['temp_max'],df1['wind'],color='green')
plt.scatter(df2['temp_max'],df2['wind'],color='red')
plt.scatter(df3['temp_max'],df3['wind'],color='blue')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Temp Max')
```



Lembar Kerja Mahasiswa

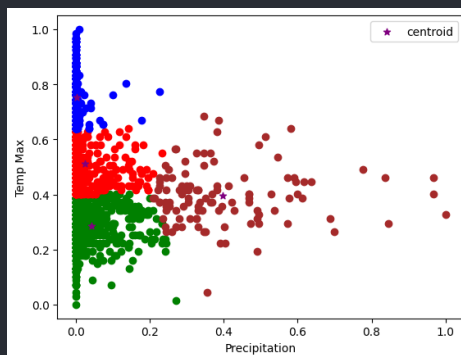
Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 10/15

```
plt.ylabel('Wind')  
plt.legend()
```



```
#Cluster 4
```

```
kmean = KMeans(n_clusters=4)  
y_predicted = kmean.fit_predict(df[['precipitation', 'temp_max']])  
y_predicted  
  
df['cluster'] = y_predicted  
df.head()  
kmean.cluster_centers_  
  
df1 = df[df.cluster==0]  
df2 = df[df.cluster==1]  
df3 = df[df.cluster==2]  
df4 = df[df.cluster==3]  
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='brown')  
plt.scatter(kmean.cluster_centers_[0,0], kmean.cluster_centers_[0,1], color='purple',  
marker='*', Label='centroid')  
plt.xlabel('Precipitation')  
plt.ylabel('Temp Max')  
plt.legend()
```





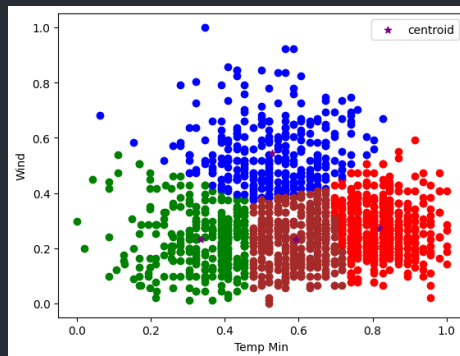
Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 11/15

```
kmean = KMeans(n_clusters=4)
y_predicted = kmean.fit_predict(df[['temp_min', 'wind']])
y_predicted

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

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
df4 = df[df.cluster==3]
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')
plt.scatter(df4['temp_min'], df4['wind'], color='brown')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
            marker='*', label='centroid')
plt.xlabel('Temp Min')
plt.ylabel('Wind')
plt.legend()
```



```
kmean = KMeans(n_clusters=4)
y_predicted = kmean.fit_predict(df[['precipitation', 'temp_min']])
y_predicted

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

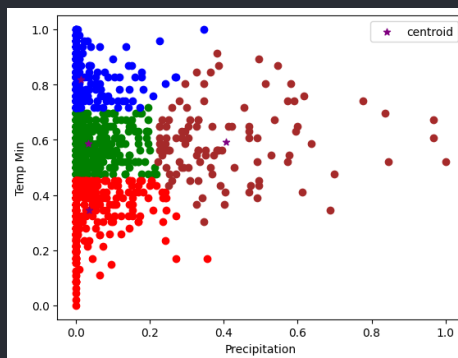
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 12/15

```
df4 = df[df.cluster==3]
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='brown')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Precipitation')
plt.ylabel('Temp Min')
plt.legend()
```



```
kmean = KMeans(n_clusters=4)
y_predicted = kmean.fit_predict(df[['temp_max', 'wind']])
y_predicted

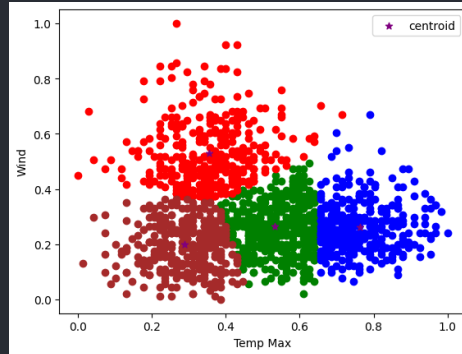
df['cluster'] = y_predicted
df.head()
kmean.cluster_centers_

df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
df4 = df[df.cluster==3]
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='brown')
plt.scatter(kmean.cluster_centers_[0], kmean.cluster_centers_[1], color='purple',
marker='*', Label='centroid')
plt.xlabel('Temp Max')
plt.ylabel('Wind')
plt.legend()
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 13/15



6. Analisislah perbandingan scatter plot sebelum clustering dan setelah clustering

Menurut saya perbedaan yang terjadi antar scatter plot sebelum clustering dan setelah clustering adalah Dimana pada saat kita belum melakukan clustering tampak hanya terlihat seperti menggambarkan 1 pola saja dengan 1 warna, setelah kita melakukan clustering bisa kita lihat pada jawaban saya diatas adalah Dimana dalam plot yang terjadi ini ternyata ada lebih banyak pola yang bisa di gambarkan, ternyata masih banyak ciri ciri pada setiap datanya yang akhirnya terlihat setelah pembagian sesuai dengan cluster-nya.

Yang mana pada soal diatas kita diminta untuk melakukan clustering sebanyak 4 bagian pada setiap atribut yang ditentukan. Dijelaskan bahwa pada code yang saya cantumkan variable kmean yang berisikan jumlah dari banyaknya cluster jika $kmean = 2$ maka akan terbagi menjadi 2 warna begitu terus sampai ke $kmean = 4$

Karena adanya clustering ini, saya dapat memahami dan tahu tentang ciri ciri dari setiap cluster yang telah dibagi sesuai dengan hasil learningnya..

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

Menurut saya secara visual sendiri kita dapat memahami dengan lebih mudah clustering 4, karena semakin banyak centroid yang di bawa oleh cluster tersebut lebih mudah kita menganalisis perbedaannya.



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 14/15

HASIL DAN ANALISIS DATA

- Hasil analisis data pada LKM 5 :
 1. Melakukan normalisasi min-Max terhadap dataset yang sudah disediakan (seattle-weather.csv) yang bisa di dapatkan melalui google drive. Yang mana tujuan dari normalisasi ini ada mengubah nilai angka di dalam dataset menjadi float.
 2. Melakukan visualisasi scatter plot pada atribut precipitation, temp_min, temp_max, dan wind sebelum dilakukannya clustering
 3. Melakukan clustering dengan nilai 2,3,4
 4. Mencari nilai yang paling efektif pada perhitungan analisis Elbow Method, dengan melakukan perhitungan dari range 2-11 terjadi penurunan inersia yang drastis, dan hal itu berakhir pada nilai $X = 4$, jadi menurut saya yang terbaik pada nilai $X = 4$
 5. Melakukan visualisasi scatter yang sudah kita lakukan clustering,
 6. Melakukan analisis perbedaan, antara plot yang sebelum terjadinya clustering dan sesudah terjadinya clustering disemua iterasi yang kita lakukan yaitu kmean = 1 sampai dengan kmean = 4
 7. Mencari nilai kmean terbaik dengan melakukan analisis secara visual yang mana menurut saya hasil terbaik terjadi pada saat kmean = 4

KESIMPULAN

Kesimpulan LKM 5 kali ini kita dapat mengetahui Bagaimana cara menggunakan library library python lainnya, yaitu yang kita gunakan kali ini adalah K-Means, Clustering. Kita juga dapat mengetahui perubahan yang terjadi pada saat kita melakukan clustering terhadap sebuah dataset. Kemudian kita juga dapat mengetahui apa itu Elbow Method yang mana Elbow Method ini adalah untuk menentukan jumlah kluster optimal pada sebuah dataset dengan melihat titik inersia sehingga membentuk bentuk siku pada visualisasi plot. Ini akan menunjukkan jumlah kluster optimal yang digunakan dalam dataset.

Link Google Colab	https://colab.research.google.com/drive/1jwGYLwKtcABpySr2oM5H7bWwG6ID967i?usp=sharing
Link Youtube (Unlisted)	https://youtu.be/bGHCiD_3AXQ

Jember,2024

Mengetahui,
Dosen Datamining

Asisten,



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 15/15

Fajrin Nurman Arifin, S.T., M.Eng
NIP. 198511282015041002

(Nama Jelas)
NIM.