



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 1/11

NIM	222410101034
Nama	Akmal Ihab Syauqi
Kelas	A
Program Studi	Sistem Informasi
Asisten	1. Renata Sayidatul Arikha (212410101057) 2. Aprodhita Nanda Eka Wijaya (212410101071)

LANGKAH KERJA

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

1. Normalisasikan dataset seattle-weather menggunakan normalisasi minmax

```
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 : 2/11

Hasil:

	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

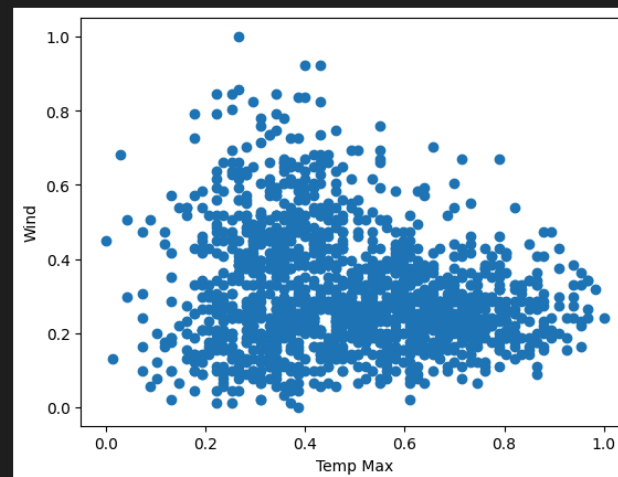
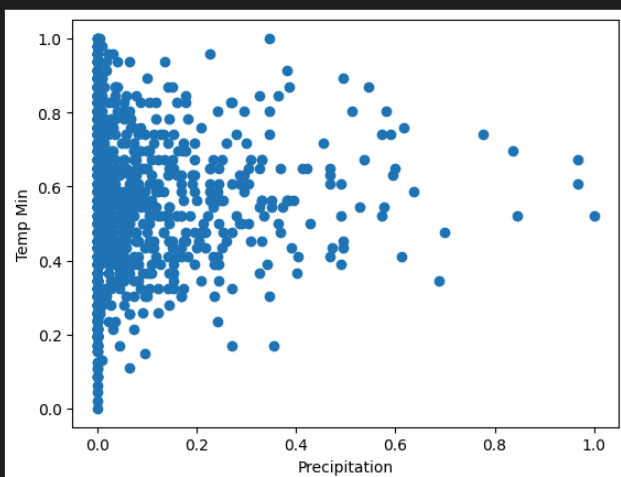
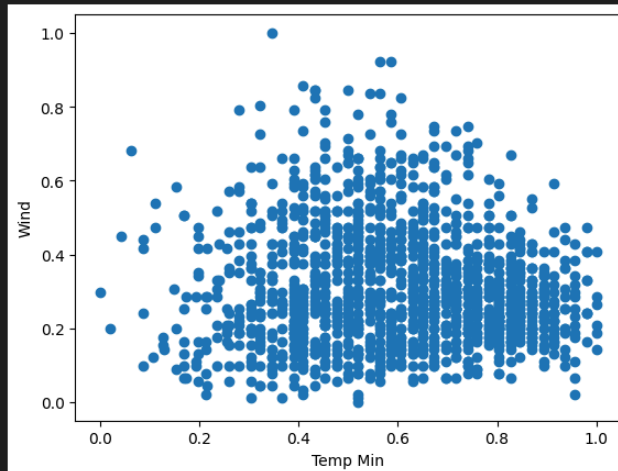
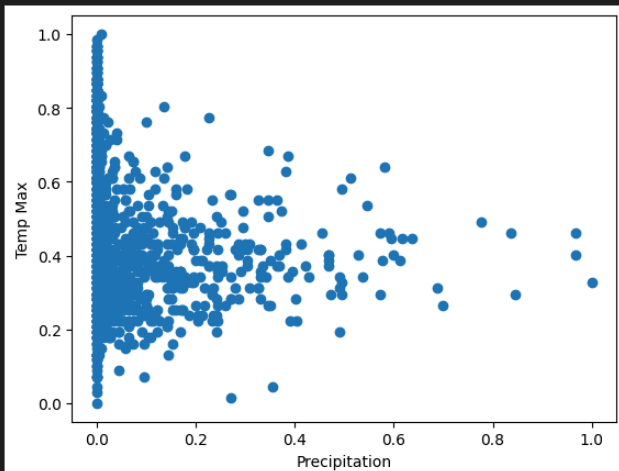
```
plt.scatter(df['precipitation'],df['temp_max'])  
plt.xlabel('Precipitation')  
plt.ylabel('Temp Max')  
  
plt.scatter(df['temp_min'],df['wind'])  
plt.xlabel('Temp Min')  
plt.ylabel('Wind')  
  
plt.scatter(df['precipitation'],df['temp_min'])  
plt.xlabel('Precipitation')  
plt.ylabel('Temp Min')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

Hasil:



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 3/11



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']])
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 4/11

```
# 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

```
inertia = []

X = df[['precipitation', 'temp_max']]
# X = df[['temp_min', 'wind']]
# X = df[['precipitation', 'temp_min']]
# X = df[['temp_max', 'wind']]

K = range(1, 11)

for k in K:
    clustering = KMeans(n_clusters=k, random_state=42)
    clustering.fit(X)
    inertia.append(clustering.inertia_)

plt.plot(K, inertia, 'o-')
plt.xlabel('Values of K')
```

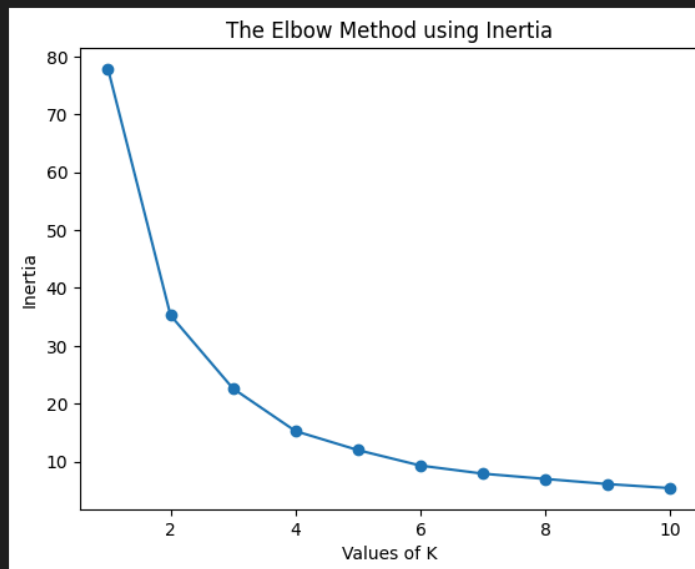


Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 5/11

```
plt.ylabel('Inertia')  
plt.title('The Elbow Method using Inertia')  
plt.show()
```

Hasil:



Berdasarkan perhitungan Elbow Method berikut, yang merupakan Nilai K terbaik adalah K-Means = 4

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')
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 6/11

```
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')
# 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')
```

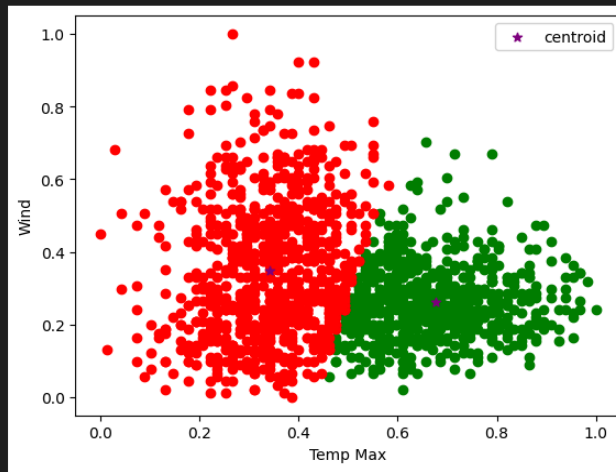
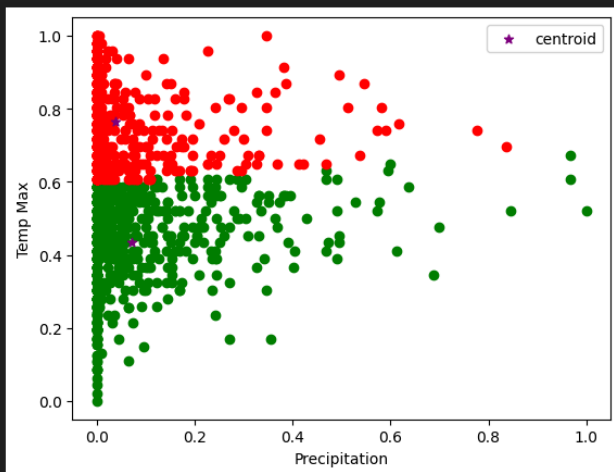
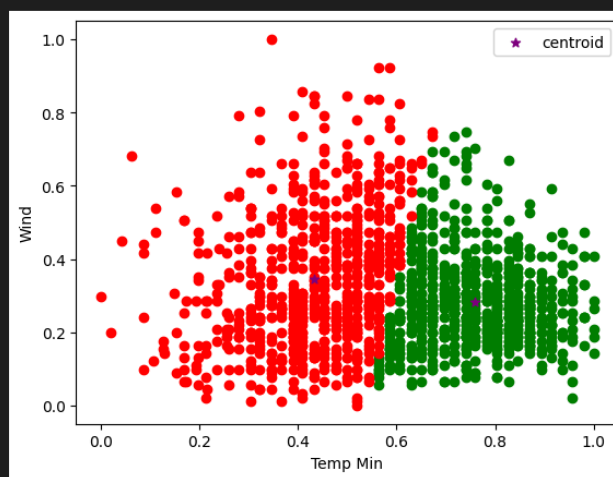
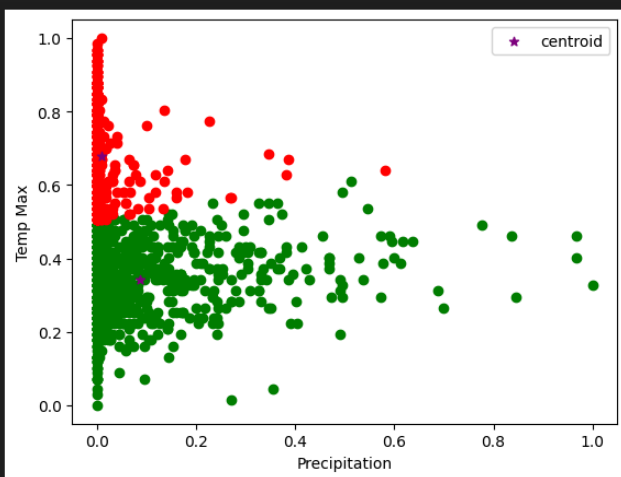


Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 7/11

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

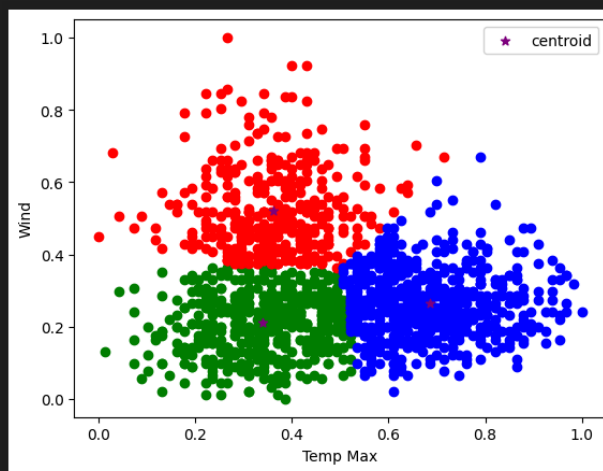
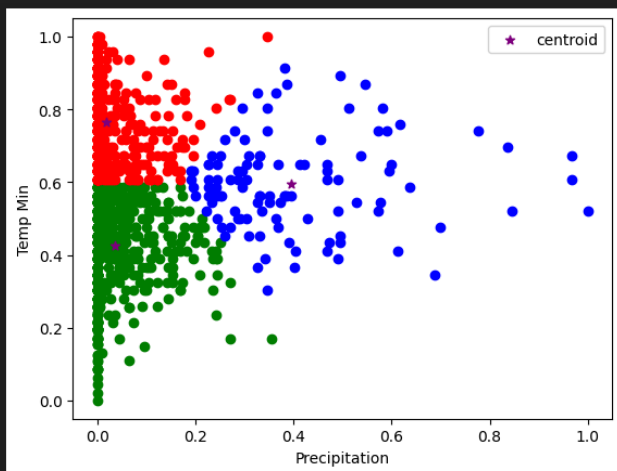
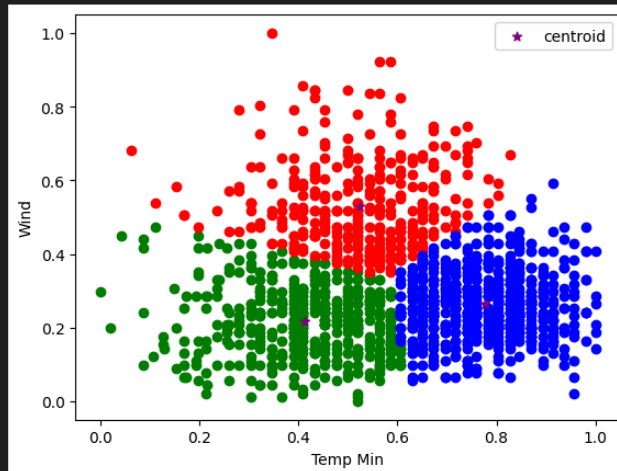
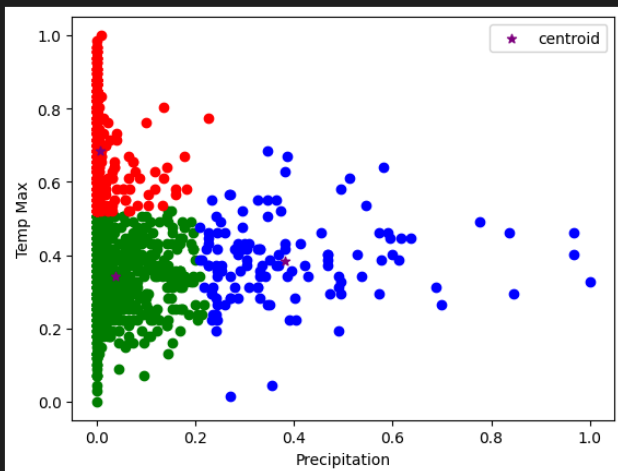
Hasil:





Lembar Kerja Mahasiswa

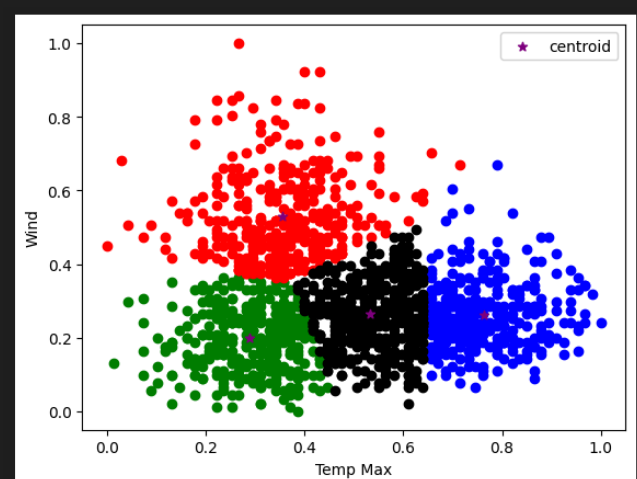
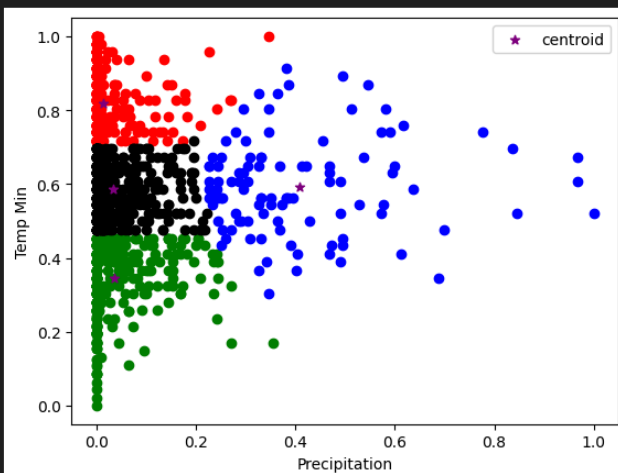
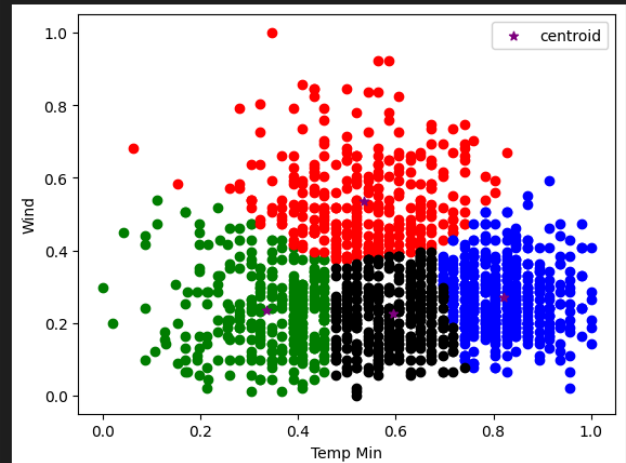
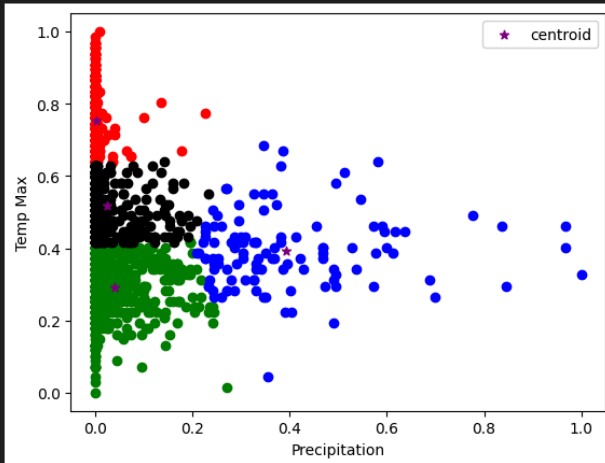
Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 8/11





Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : K-Means Clustering
Halaman : 9/11



6. Analisislah perbandingan scatter plot sebelum clustering dan setelah clustering

- Berdasarkan output scatter plot yang sudah ditampilkan, perbedaan yang terlihat antara sebelum dan sesudah dilakukan clustering adalah Output scatter plot sebelum diclustering hanya menggambarkan satu cluster saja, dengan warna biru, namun setelah dilakukan clustering dan dicari centroid-nya, scatter plot dari data seattle-weather akan dibagi menjadi beberapa bagian sesuai dengan banyaknya K atau banyaknya cluster yang akan dikelompokkan
- Jika $K = 2$ maka akan dibagi menjadi 2 cluster dengan 2 warna, $K = 3$ menjadi 3 cluster, dan $K = 4$ menjadi 4 cluster.



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
K-Means Clustering
Bahasan :
Halaman : 10/11

- Dengan adanya clustering ini, kita jadi lebih paham dan tahu karakteristik dari setiap cluster, dan tidak ada cluster yang masuk dalam kategori ambigu/memiliki ciri dari clusternya sendiri dan cluster lain.

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

- Jika kita lakukan analisis secara visual maka dapat dilihat bahwa K terbaik disini adalah $K = 4$, karena dari perhitungan menggunakan Elbow Method juga ditunjukkan penurunan inersia drastis hanya sampai di $K = 4$.
- Kemudian dilihat dari scatter plot-nya bahwa pada $K = 4$ struktur plotnya lebih nyaman dibaca dan juga kita bisa tahu lebih pas tentang karakteristik setiap clusternya, tidak terlalu ambigu, tidak terlalu spesifik juga.

HASIL DAN ANALISIS DATA

- Hasil dan analisis data pada LKM 5 kali ini adalah sebagai berikut:
 1. Melakukan normalisasi Min-Max terhadap dataset seattle-weather.csv yang didapat dari google drive praktikum, normalisasi Min-Max bertujuan untuk mengubah nilai-nilai angka di dalam dataset menjadi float dengan range dari 0 sampai 1 untuk memudahkan clustering nantinya.
 2. Kemudian kita mencoba untuk menampilkan scatter plot dari setiap perbandingan antara atribut precipitation, temp_min, temp_max, dan wind sebelum melakukan clustering, hasilnya tampil plot dengan titik dari setiap isi dari atribut dataset yang dibandingkan.
 3. Melakukan clustering terhadap dataset tersebut, dengan K sebesar 2, 3, dan 4
 4. Kita mencari Nilai K yang paling efektif untuk digunakan, menggunakan perhitungan Elbow Method yang menghitung inersia dari $K = 1$ hingga $K = n$. dan ditemukan bahwa puncak penurunan inersia drastis, berakhir di $K = 4$.
 5. Kemudian kita tampilkan dari setiap scatter plot dengan nilai K yang berbeda, dan melihat perbedaannya
 6. Kita menganalisis perbedaan, antara scatter plot data yang belum dilakukan clustering, dengan yang sudah dilakukan clustering, baik itu $K = 2$, $K = 3$, atau $K = 4$.



Lembar Kerja Mahasiswa

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

7. Kemudian kita mencari nilai K terbaik dengan melakukan analisis secara visual terhadap output-output dari scatter plot yang sudah ditampilkan, dan yang terbaik adalah $K = 4$.

KESIMPULAN

- Kesimpulan dari pengerjaan LKM 5 ini adalah, kita dapat mengetahui bagaimana cara kerja K-Means Clustering, dan bagaimana dampaknya terhadap visualisasi data, kita juga mengetahui bagaimana cara kita menambahkan nilai K dan mengaturnya supaya menciptakan cluster sesuai dengan centroid yang sudah ditentukan secara otomatis, kita juga tahu fungsi dari Elbow Method yaitu untuk mencari titik drastis inersia terakhir dalam menentukan Nilai K terbaik.

Link Google Colab

 LKM 5 K-Means Clustering_1034.ipynb

Link Youtube (Unlisted)

<https://youtu.be/eay2BboeO1A>

Jember,2024

Mengetahui,
Dosen Datamining

Asisten,

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

(Nama Jelas)
NIM.