

ANALISIS CLUSTER CURAH HUJAN PROVINSI JAWA TENGAH BULAN JANUARI TAHUN 2015 - 2024 DENGAN ALGORITMA K-MEANS

Rifaldi Achmad Faisal

**BBMKG Region II Research Project
Supervisor : Ismi Amalia S.Si**



DAFTAR ISI

03 LATAR
BELAKANG

04 TUJUAN

05 METODE
PENELITIAN

09 ANALISIS
DATA
EKSPLORATIF

16 K - MEANS
CLUSTERING

27 DETAIL
TIAP
CLUSTER

32 KESIMPULAN

33 DAFTAR
PUSTAKA

LATAR BELAKANG

- Indonesia merupakan negara beriklim tropis



- Jawa Tengah salah satu provinsi terluas di Pulau Jawa



- Curah hujan yang beragam di Provinsi Jawa Tengah



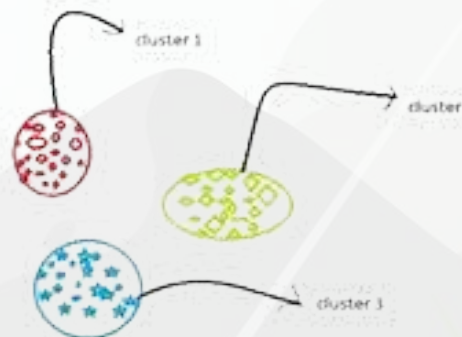
- Jawa Tengah memiliki enam stasiun BMKG yang tersebar.



- Terdapat suatu algoritma pengelompokan bernama K-Means Clustering.



- Beberapa daerah di Jawa Tengah dikelompokkan berdasarkan kemiripan tingkat curah hujan menggunakan K-Means Clustering.



TUJUAN

1. Menentukan berapa banyak cluster optimal yang terbentuk akibat adanya perbedaan curah hujan di Provinsi Jawa Tengah.
2. Mengetahui proses kerja algoritma K-Means Clustering.
3. Menentukan cluster pada tiap stasiun di Provinsi Jawa Tengah.
4. Mengetahui letak persebaran cluster beserta informasi berdasarkan statistika deskriptif, letak geografis dan topografisnya.

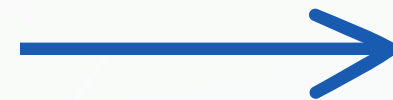


METODE PENELITIAN

APA ITU K-MEANS CLUSTERING?

Artificial Intelligence (AI):

Cabang ilmu komputer yang bertujuan untuk membuat program komputer mampu melakukan tugas yang biasanya membutuhkan kecerdasan manusia.



Machine Learning (ML):

Cabang Artificial Intelligence yang memungkinkan komputer untuk belajar dari data tanpa perlu diprogram secara eksplisit.



Algoritma K-Means Clustering:

Metode yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan fitur atau atributnya.



Unsupervised Learning:

Cabang dari Machine Learning di mana komputer belajar dari data tanpa petunjuk atau label yang diberikan.

EUCLIDIAN DISTANCE PADA K-MEANS

$$dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \dots$$

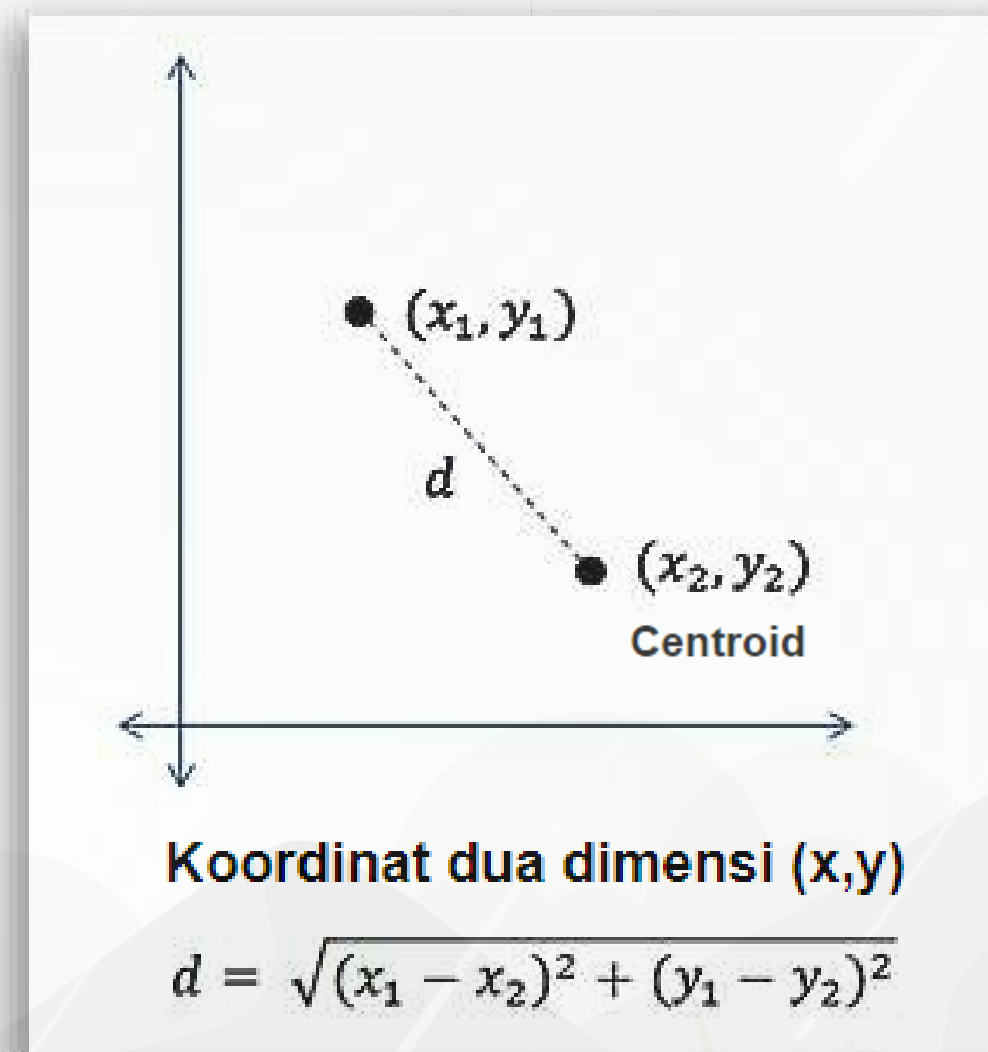
Keterangan:

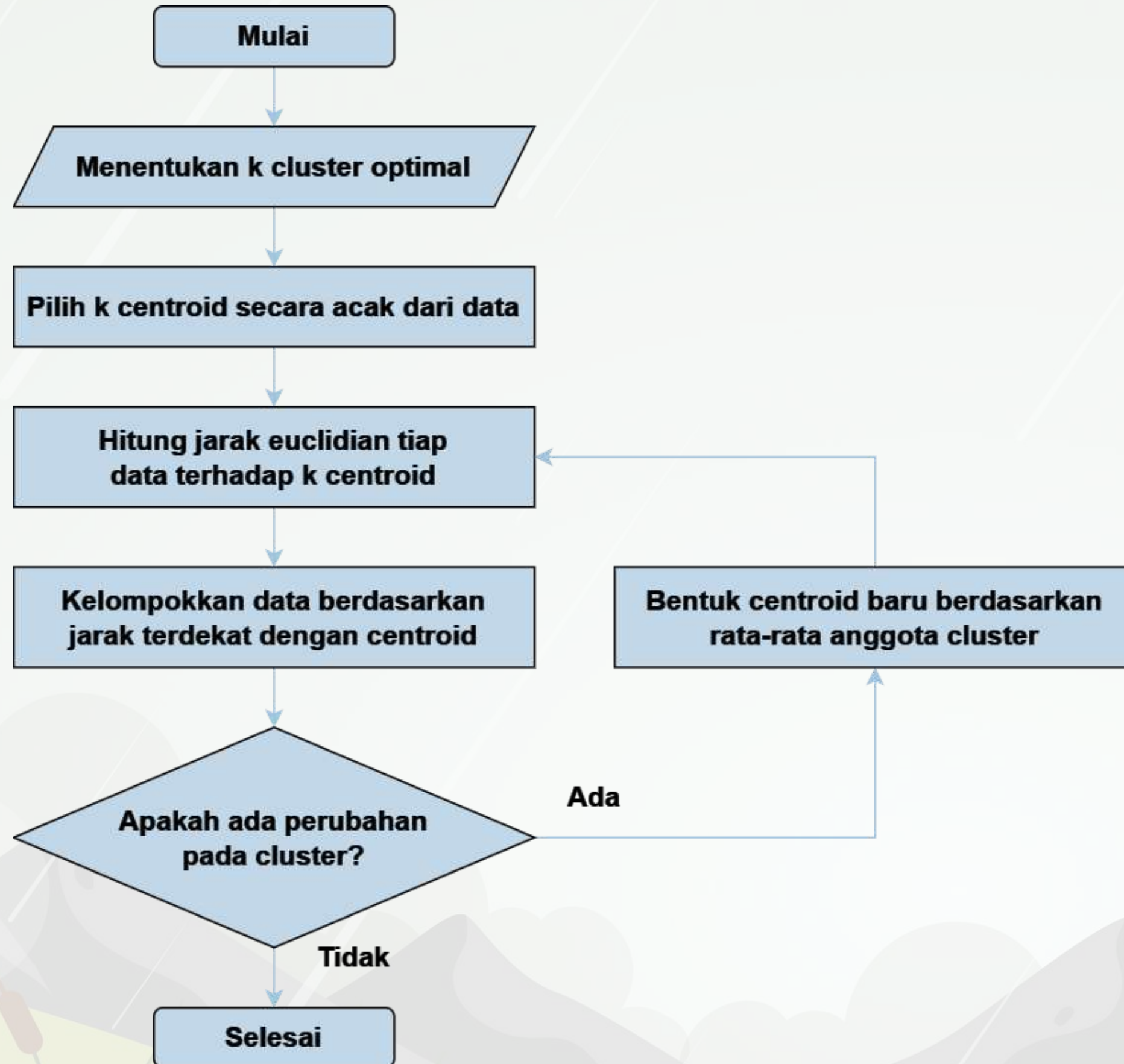
dist : Jarak dari titik pada data ke centroid (pusat)

$(x_1 - x_2)$: Jarak koordinat x_1 ke x_2

$(y_1 - y_2)$: Jarak koordinat y_1 ke y_2

$(z_1 - z_2)$: Jarak koordinat z_1 ke z_2





FLOW CHART ALGORITMA K-MEANS CLUSTERING

ANALISIS DATA EKSPLORATIF



MENGENAL DATA



Sumber Data:

Data Curah Hujan (mm) bulanan di enam stasiun di Provinsi Jawa Tengah pada bulan Januari tahun 2015 - 2024



Nama Stasiun:

- Stamet Tunggul Wulung ▼
- Stageof Banjarnegara ▼
- Stamet Maritim Tegal ▼
- Staklim Jawa Tengah ▼
- Stamet Maritim Tanjung Emas ▼
- Stamet Ahmad Yani ▼

IMPORT LIBRARY DAN FILE DATA

```
[1] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
```

```
[3] dff = pd.read_excel('/content/Jawa Tengah - Curah Hujan Januari.xlsx')
dxx = pd.read_excel('/content/Curah Hujan Tahunan.xlsx')
df = dff.drop(columns=['Stasiun'])
dxx.set_index('Tahun', inplace=True)
dff.set_index('Stasiun', inplace=True)
dff
```

- Import beberapa library yang dibutuhkan dan juga file excel ke dalam notebook.
- Data utama ada pada '/content/Jawa Tengah - Curah Hujan Januari.xlsx'.

DATA PREPROCESSING

Detail:

- Setengah data berupa data harian (setengah sisanya sudah data bulanan)
- Mengubah nilai 8888 ke 0 pada data harian
- Menjumlahkan data harian hingga mendapat data bulanan
- Mengubah kolom Stasiun menjadi index

| Stasiun | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Stamet Tunggul Wulung | 294.0 | 171.9 | 371.0 | 298.1 | 365.0 | 157.5 | 286.2 | 364.4 | 139.2 | 444.3 |
| Stageof Banjarnegara | 460.9 | 258.2 | 480.4 | 540.1 | 765.9 | 196.5 | 615.0 | 244.7 | 76.6 | 517.4 |
| Stamet Maritim Tegal | 419.0 | 234.0 | 396.0 | 85.0 | 551.0 | 524.2 | 235.7 | 361.8 | 260.7 | 231.3 |
| Staklim Jawa Tengah | 238.2 | 269.0 | 282.0 | 348.6 | 217.0 | 301.3 | 272.7 | 329.3 | 297.7 | 303.4 |
| Stamet Maritim Tanjung Emas | 215.0 | 222.0 | 247.3 | 268.8 | 370.0 | 336.3 | 300.8 | 206.0 | 314.9 | 319.1 |
| Stamet Ahmad Yani | 221.2 | 284.4 | 256.4 | 350.7 | 234.8 | 321.1 | 215.4 | 266.2 | 343.5 | 203.6 |

STATISTIKA DESKRIPTIF

```
[4] summary_stats = dxx.describe()
summary_stats = summary_stats.drop(['25%', '50%', '75%'])
summary_stats_int = summary_stats.astype(int)
summary_stats_int
```

| Index | Stamet Tunggal Wulung | Stageof Banjarnegara | Stamet Maritim Tegal | Staklim Jawa Tengah | Stamet Maritim Tanjung Emas | Stamet Ahmad Yani |
|-------|--------------------------|-------------------------|-------------------------|------------------------|--------------------------------|----------------------|
| count | 10 | 10 | 10 | 10 | 10 | 10 |
| mean | 289 | 415 | 329 | 285 | 280 | 269 |
| std | 103 | 213 | 146 | 39 | 56 | 53 |
| min | 139 | 76 | 85 | 217 | 206 | 203 |
| max | 444 | 765 | 551 | 348 | 370 | 350 |

Detail:

- Count mewakili banyaknya tahun pada tiap stasiun
- Mean adalah rata-rata tiap stasiun
- Std adalah standar deviasi tiap stasiun
- Min adalah nilai terkecil tiap stasiun
- Max adalah nilai terbesar tiap stasiun

PRINCIPAL COMPONENT ANALYSIS

```
[6] # PCA
pca = PCA(n_components=2)
dataset_pca = pca.fit_transform(df)

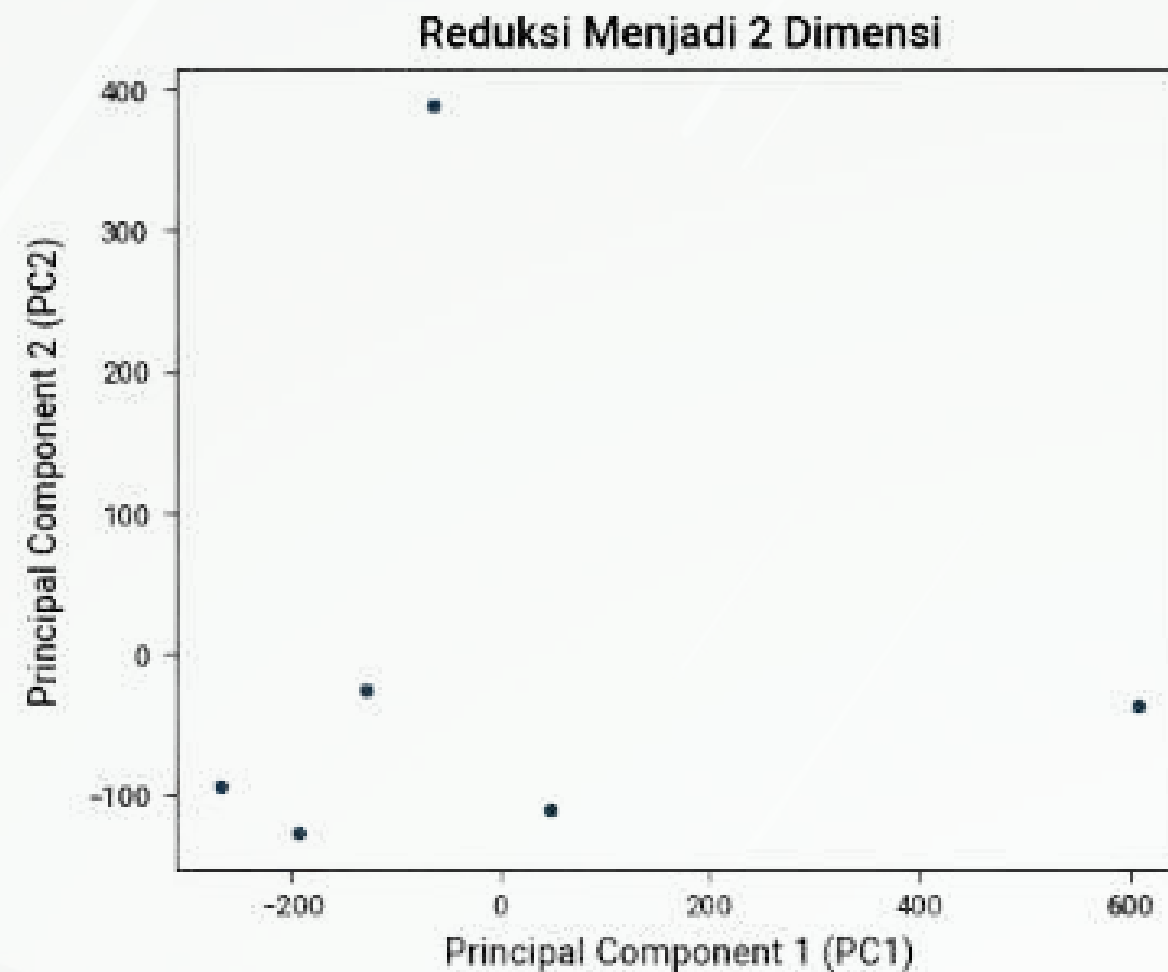
# Buat DataFrame baru dari hasil PCA
dfPCA = pd.DataFrame(data=dataset_pca, columns=['PC1', 'PC2'])

# Tampilkan hasil PCA
print("Data setelah PCA:")
print(dfPCA)

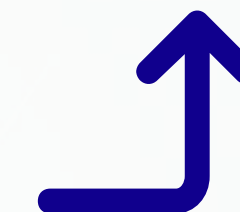
# Visualisasi hasil PCA
plt.figure(figsize=(10, 6))
plt.scatter(dfPCA['PC1'], dfPCA['PC2'])
plt.title('Reduksi Menjadi 2 Dimensi')
plt.xlabel('Principal Component 1 (PC1)')
plt.ylabel('Principal Component 2 (PC2)')
plt.show()
```

- Menggunakan principle component analysis (PCA) untuk mereduksi dimensi data menjadi dua dimensi.
- Ini digunakan untuk memudahkan ketika melakukan visualisasi yang berkaitan dengan analisis k-means.

PRINCIPAL COMPONENT ANALYSIS



| Stasiun | PC1 | PC 2 |
|-----------------------------|-------------|-------------|
| Stamet Tunggul Wulung | 46.140764 | -109.451707 |
| Stageof Banjarnegara | 607.933335 | -36.223451 |
| Stamet Maritim Tegal | -65.123490 | 389.940428 |
| Staklim Jawa Tengah | -192.736440 | -125.908602 |
| Stamet Maritim Tanjung Emas | -129.297340 | -24.652412 |
| Stamet Ahmad Yani | -266.916829 | -93.704256 |



REDUKSI MENJADI 2 VARIABEL

| Stasiun | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Stamet Tunggul Wulung | 294.0 | 171.9 | 371.0 | 298.1 | 365.0 | 157.5 | 286.2 | 364.4 | 139.2 | 444.3 |
| Stageof Banjarnegara | 460.9 | 258.2 | 480.4 | 540.1 | 765.9 | 196.5 | 615.0 | 244.7 | 76.6 | 517.4 |
| Stamet Maritim Tegal | 419.0 | 234.0 | 396.0 | 85.0 | 551.0 | 524.2 | 235.7 | 361.8 | 260.7 | 231.3 |
| Staklim Jawa Tengah | 238.2 | 269.0 | 282.0 | 348.6 | 217.0 | 301.3 | 272.7 | 329.3 | 297.7 | 303.4 |
| Stamet Maritim Tanjung Emas | 215.0 | 222.0 | 247.3 | 268.8 | 370.0 | 336.3 | 300.8 | 206.0 | 314.9 | 319.1 |
| Stamet Ahmad Yani | 221.2 | 284.4 | 256.4 | 350.7 | 234.8 | 321.1 | 215.4 | 266.2 | 343.5 | 203.6 |



K-MEANS CLUSTERING

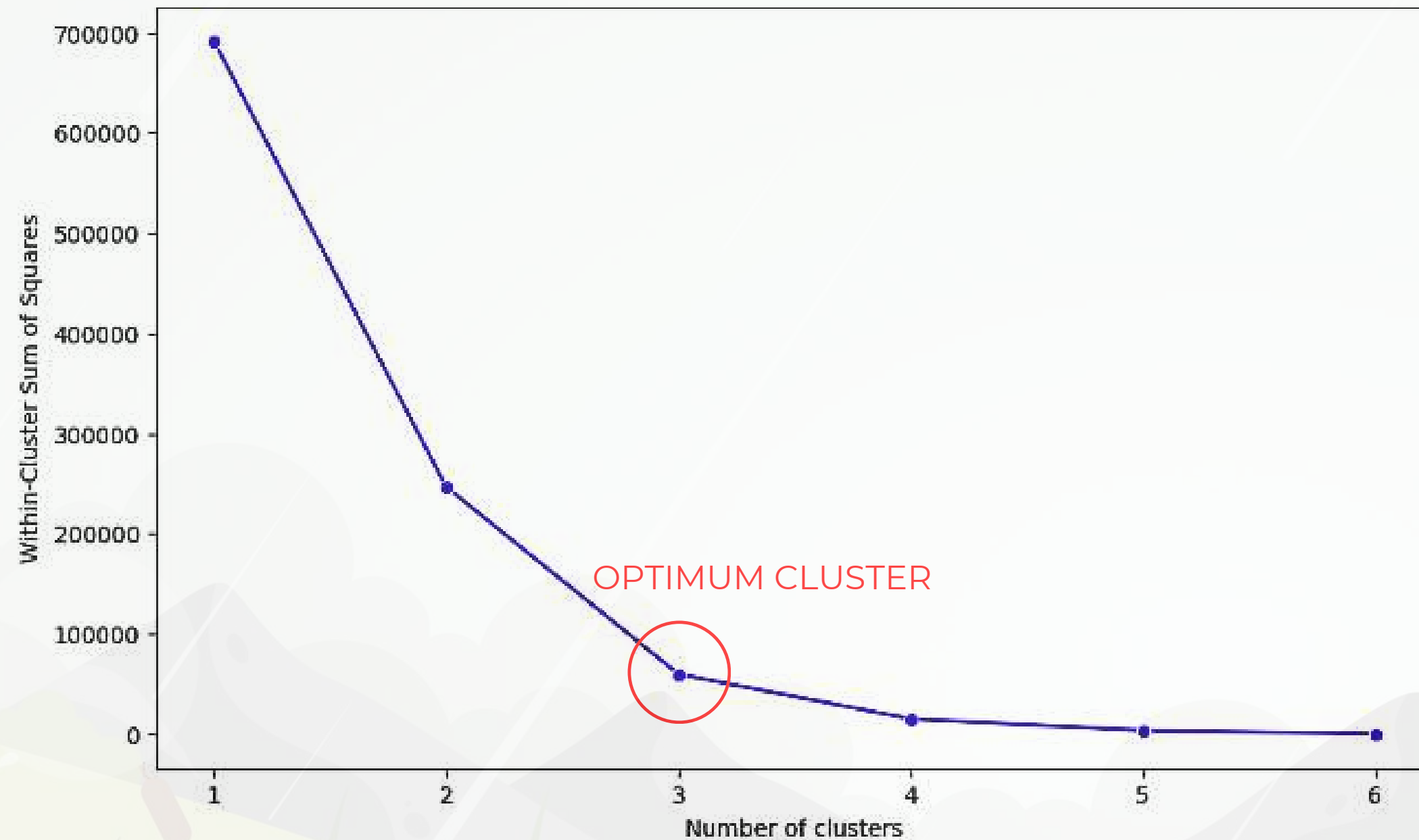
ELBOW METHOD

```
[7] wcss = [] # WCSS: Within-Cluster Sum of Squares
    for i in range(1,7):
        kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
        kmeans.fit(dfPCA)
        wcss.append(kmeans.inertia_)
```

```
[8] plt.figure(figsize=(10, 6))
    sns.lineplot(x=range(1, 7), y=wcss, marker='o', color='blue')
    plt.title('The Elbow Method')
    plt.xlabel('Number of clusters')
    plt.ylabel('Within-Cluster Sum of Squares')
    plt.show()
```

- WCSS (Within Cluster Sum of Squares) digunakan untuk menghitung jarak dari masing-masing data ke centroid lalu dijumlah dan dikuadratkan.
- Jika nilai WCSS terlalu besar, maka centroid baru muncul di tengah data. Begitu seterusnya sampai nilai WCSS berhenti turun secara signifikan.
- WCSS divisualisasikan menggunakan Elbow Method untuk mencari jumlah cluster paling optimal.

ELBOW METHOD



Detail:

- Cluster paling optimal sebanyak 3
- Jika terlalu banyak akan menyebabkan overfitting
- Jika terlalu sedikit representasinya tidak akan detail

AMBIL SEBARANG CENTROID

Note : Perhitungan manual menggunakan Ms. Excel

| Stasiun | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Stamet Tunggul Wulung | 294.0 | 171.9 | 371.0 | 298.1 | 365.0 | 157.5 | 286.2 | 364.4 | 139.2 | 444.3 |
| Stageof Banjarnegara | 460.9 | 258.2 | 480.4 | 540.1 | 765.9 | 196.5 | 615.0 | 244.7 | 76.6 | 517.4 |
| Stamet Maritim Tegal | 419.0 | 234.0 | 396.0 | 85.0 | 551.0 | 524.2 | 235.7 | 361.8 | 260.7 | 231.3 |
| Staklim Jawa Tengah | 238.2 | 269.0 | 282.0 | 348.6 | 217.0 | 301.3 | 272.7 | 329.3 | 297.7 | 303.4 |
| Stamet Maritim Tanjung Emas | 215.0 | 222.0 | 247.3 | 268.8 | 370.0 | 336.3 | 300.8 | 206.0 | 314.9 | 319.1 |
| Stamet Ahmad Yani | 221.2 | 284.4 | 256.4 | 350.7 | 234.8 | 321.1 | 215.4 | 266.2 | 343.5 | 203.6 |

CLUSTER 0

CLUSTER 1

CLUSTER 2

ITERASI KE-1

Note : Perhitungan manual menggunakan Ms. Excel

| Stasiun | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Stamet Tunggul Wulung | 294.0 | 171.9 | 371.0 | 298.1 | 365.0 | 157.5 | 286.2 | 364.4 | 139.2 | 444.3 |
| Stageof Banjarnegara | 460.9 | 258.2 | 480.4 | 540.1 | 765.9 | 196.5 | 615.0 | 244.7 | 76.6 | 517.4 |
| Stamet Maritim Tegal | 419.0 | 234.0 | 396.0 | 85.0 | 551.0 | 524.2 | 235.7 | 361.8 | 260.7 | 231.3 |
| Staklim Jawa Tengah | 238.2 | 269.0 | 282.0 | 348.6 | 217.0 | 301.3 | 272.7 | 329.3 | 297.7 | 303.4 |
| Stamet Maritim Tanjung Emas | 215.0 | 222.0 | 247.3 | 268.8 | 370.0 | 336.3 | 300.8 | 206.0 | 314.9 | 319.1 |
| Stamet Ahmad Yani | 221.2 | 284.4 | 256.4 | 350.7 | 234.8 | 321.1 | 215.4 | 266.2 | 343.5 | 203.6 |

Jarak ke centroid 0 :

$$\begin{aligned} &\sqrt{(294 - 238.2)^2 + (171.9 - 222)^2 + \dots + (444.3 - 444.3)^2} \\ &\sqrt{(460.9 - 238.2)^2 + (258.2 - 222)^2 + \dots + (517.4 - 444.3)^2} \\ &\sqrt{(419 - 238.2)^2 + (234 - 222)^2 + \dots + (231.3 - 444.3)^2} \\ &\sqrt{(238.2 - 238.2)^2 + (282 - 222)^2 + \dots + (303.4 - 444.3)^2} \\ &\sqrt{(215 - 238.2)^2 + (222 - 222)^2 + \dots + (319.1 - 444.3)^2} \\ &\sqrt{(221.2 - 238.2)^2 + (284.4 - 222)^2 + \dots + (203.6 - 444.3)^2} \end{aligned}$$

| Centroid | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|----------|-------|------|------|-------|-------|-------|-------|-------|-------|-------|
| 0 | 238,2 | 222 | 282 | 298,1 | 365 | 157,5 | 286,2 | 364,4 | 139,2 | 444,3 |
| 1 | 460,9 | 234 | 371 | 540,1 | 551 | 524,2 | 235,7 | 244,7 | 76,6 | 517,4 |
| 2 | 221,2 | 269 | 396 | 85 | 765,9 | 336,3 | 272,7 | 361,8 | 260,7 | 231,3 |

Note :

Karena perhitungan jarak ke setiap centroid sama, maka yang dipakai dalam perhitungan ini hanya jarak masing-masing titik ke centroid 0. Dengan tujuan untuk mewakili perhitungan jarak ke centroid 1 dan centroid 2.

HASIL ITERASI KE-1

Note : Perhitungan manual menggunakan Ms. Excel

| Stasiun | C0 | C1 | C2 | Jarak Min | Cluster |
|-----------------------------|-------------|-------------|-------------|-------------|---------|
| Stamet Tunggal Wulung | 116,3814848 | 534,3048007 | 560,1565585 | 116,3814848 | 0 |
| Stageof Banjarnegara | 665,4126539 | 556,7690634 | 733,4738305 | 556,7690634 | 1 |
| Stamet Maritim Tegal | 568,325083 | 582,1984627 | 351,0089173 | 351,0089173 | 2 |
| Staklim Jawa Tengah | 306,1552057 | 599,8844222 | 626,8251989 | 306,1552057 | 0 |
| Stamet Maritim Tanjung Emas | 326,2583486 | 565,1805729 | 500,5877645 | 326,2583486 | 0 |
| Stamet Ahmad Yani | 407,0045946 | 648,7019423 | 626,6305451 | 407,0045946 | 0 |

Detail:

- Nilai C0, C1, dan C2 merupakan hasil perhitungan jarak euclidean.
- Kolom "Jarak Minimum" mengambil nilai jarak terkecil dari ketiga centroid.
- Kolom "Cluster" memberi label cluster sesuai jarak minimum.

CENTROID BARU

Note : Perhitungan manual menggunakan Ms. Excel

| Stasiun | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | Cluster |
|-----------------------------|-------|---------|---------|--------|-------|--------|---------|---------|---------|-------|---------|
| Stamet Tunggul Wulung | 294 | 171,9 | 371 | 298,1 | 365 | 157,5 | 286,2 | 364,4 | 139,2 | 444,3 | 0 |
| Stageof Banjarnegara | 460,9 | 258,2 | 480,4 | 540,1 | 765,9 | 196,5 | 615 | 244,7 | 76,6 | 517,4 | 1 |
| Stamet Maritim Tegal | 419 | 234 | 396 | 85 | 551 | 524,2 | 235,7 | 361,8 | 260,7 | 231,3 | 2 |
| Staklim Jawa Tengah | 238,2 | 269 | 282 | 348,6 | 217 | 301,3 | 272,7 | 329,3 | 297,7 | 303,4 | 0 |
| Stamet Maritim Tanjung Emas | 215 | 222 | 247,3 | 268,8 | 370 | 336,3 | 300,8 | 206 | 314,9 | 319,1 | 0 |
| Stamet Ahmad Yani | 221,2 | 284,4 | 256,4 | 350,7 | 234,8 | 321,1 | 215,4 | 266,2 | 343,5 | 203,6 | 0 |
| Rata-Rata | 242,1 | 236,825 | 289,175 | 316,55 | 296,7 | 279,05 | 268,775 | 291,475 | 273,825 | 317,6 | |

Detail:

- Di kotak merah merupakan data yang termasuk cluster 0.
- Data cluster 0 tersebut dihitung rata-rata tiap kolom.
- Hasil rata-rata dijadikan cluster 0 yang baru.

ITERASI KE-2

Note : Perhitungan manual menggunakan Ms. Excel

| Stasiun | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Stamet Tunggul Wulung | 294.0 | 171.9 | 371.0 | 298.1 | 365.0 | 157.5 | 286.2 | 364.4 | 139.2 | 444.3 |
| Stageof Banjarnegara | 460.9 | 258.2 | 480.4 | 540.1 | 765.9 | 196.5 | 615.0 | 244.7 | 76.6 | 517.4 |
| Stamet Maritim Tegal | 419.0 | 234.0 | 396.0 | 85.0 | 551.0 | 524.2 | 235.7 | 361.8 | 260.7 | 231.3 |
| Staklim Jawa Tengah | 238.2 | 269.0 | 282.0 | 348.6 | 217.0 | 301.3 | 272.7 | 329.3 | 297.7 | 303.4 |
| Stamet Maritim Tanjung Emas | 215.0 | 222.0 | 247.3 | 268.8 | 370.0 | 336.3 | 300.8 | 206.0 | 314.9 | 319.1 |
| Stamet Ahmad Yani | 221.2 | 284.4 | 256.4 | 350.7 | 234.8 | 321.1 | 215.4 | 266.2 | 343.5 | 203.6 |

Jarak ke centroid 0 :

$$\sqrt{(294 - 242.1)^2 + (171.9 - 236.8)^2 + \dots + (444.3 - 317.6)^2}$$

$$\sqrt{(460.9 - 242.1)^2 + (258.2 - 236.8)^2 + \dots + (517.4 - 317.6)^2}$$

$$\sqrt{(419 - 242.1)^2 + (234 - 236.8)^2 + \dots + (231.3 - 317.6)^2}$$

$$\sqrt{(238.2 - 242.1)^2 + (269.0 - 236.8)^2 + \dots + (303.4 - 317.6)^2}$$

$$\sqrt{(215 - 242.1)^2 + (222 - 236.8)^2 + \dots + (319.1 - 317.6)^2}$$

$$\sqrt{(221.2 - 242.1)^2 + (284.4 - 236.8)^2 + \dots + (203.6 - 317.6)^2}$$

| Centroid | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|----------|-------|---------|---------|--------|-------|--------|---------|---------|---------|-------|
| 0 | 242,1 | 236,825 | 289,175 | 316,55 | 296,7 | 279,05 | 268,775 | 291,475 | 273,825 | 317,6 |
| 1 | 460,9 | 258,2 | 480,4 | 540,1 | 765,9 | 196,5 | 615 | 244,7 | 76,6 | 517,4 |
| 2 | 419 | 234 | 396 | 85 | 551 | 524,2 | 235,7 | 361,8 | 260,7 | 231,3 |

Note :

Karena perhitungan jarak ke setiap centroid sama, maka yang dipakai dalam perhitungan ini hanya jarak masing-masing titik ke centroid 0. Dengan tujuan untuk mewakili perhitungan jarak ke centroid 1 dan centroid 2.

HASIL ITERASI KE-2

Note : Perhitungan manual menggunakan Ms. Excel

| Stasiun | C0 | C1 | C2 | Jarak Min | Cluster | |
|-----------------------------|-------------|-------------|-------------|-------------|---------|------|
| Stamet Tunggul Wulung | 270,5223801 | 632,2795031 | 545,2257973 | 270,5223801 | 0 | SAMA |
| Stageof Banjarnegara | 750,141632 | 0 | 802,1700817 | 0 | 1 | SAMA |
| Stamet Maritim Tegal | 484,5153281 | 802,1700817 | 0 | 0 | 2 | SAMA |
| Staklim Jawa Tengah | 105,8020233 | 810,573038 | 535,3761948 | 105,8020233 | 0 | SAMA |
| Stamet Maritim Tanjung Emas | 153,815484 | 749,8843311 | 452,6363551 | 153,815484 | 0 | SAMA |
| Stamet Ahmad Yani | 178,5462073 | 878,4257055 | 538,6615635 | 178,5462073 | 0 | SAMA |

Detail:

- Nilai C0, C1, dan C2 merupakan hasil perhitungan jarak euclidean.
- Kolom "Jarak Minimum" mengambil nilai jarak terkecil dari ketiga centroid.
- Kolom "Cluster" memberi label cluster sesuai jarak minimum.
- Karena semua label cluster sama seperti sebelumnya maka iterasi konvergen.

K-MEANS CLUSTERING VISUALIZATION

```
[9] # Fitting K-Means to the dataset
kmeans = KMeans(n_clusters = 3, init = 'k-means++', random_state = 42)
y_kmeans = kmeans.fit_predict(dfPCA)

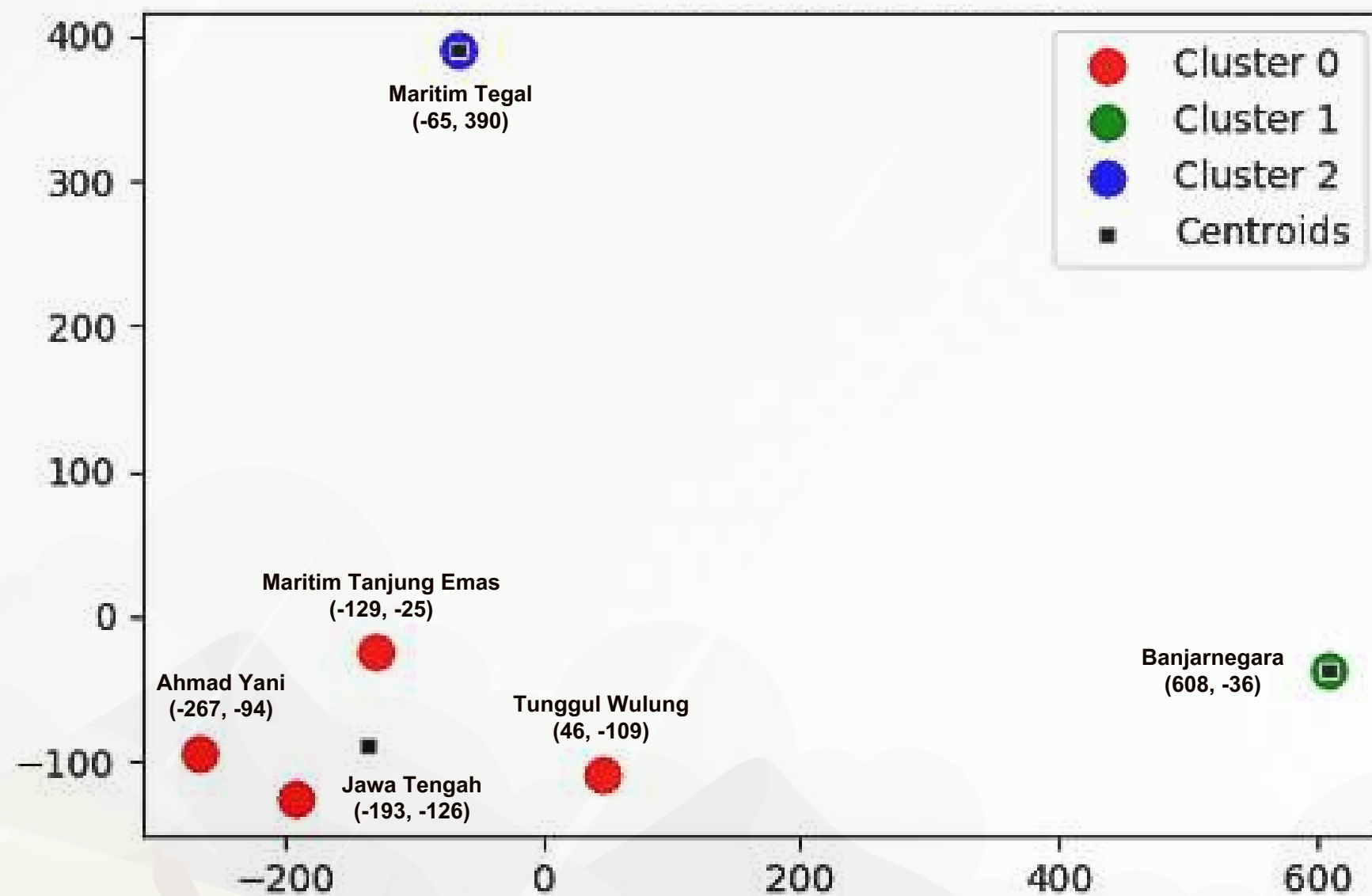
[10] plt.figure(figsize=(6, 4))

sns.scatterplot(x=dfPCA.loc[y_kmeans == 0, dfPCA.columns[0]], y=dfPCA.loc[y_kmeans == 0,
dfPCA.columns[1]], color='red', label='Cluster 0', s=120)
sns.scatterplot(x=dfPCA.loc[y_kmeans == 1, dfPCA.columns[0]], y=dfPCA.loc[y_kmeans == 1,
dfPCA.columns[1]], color='green', label='Cluster 1', s=120)
sns.scatterplot(x=dfPCA.loc[y_kmeans == 2, dfPCA.columns[0]], y=dfPCA.loc[y_kmeans == 2,
dfPCA.columns[1]], color='blue', label='Cluster 2', s=120)

sns.scatterplot(x=kmeans.cluster_centers_[0], y=kmeans.cluster_centers_[1],
color='black', label='Centroids', s=25, marker=',')

plt.grid(False)
plt.title('K-Means Clustering pada Data')
plt.xlabel('')
plt.ylabel('')
plt.legend()
plt.show()
```

K-MEANS CLUSTERING VISUALIZATION



Detail:

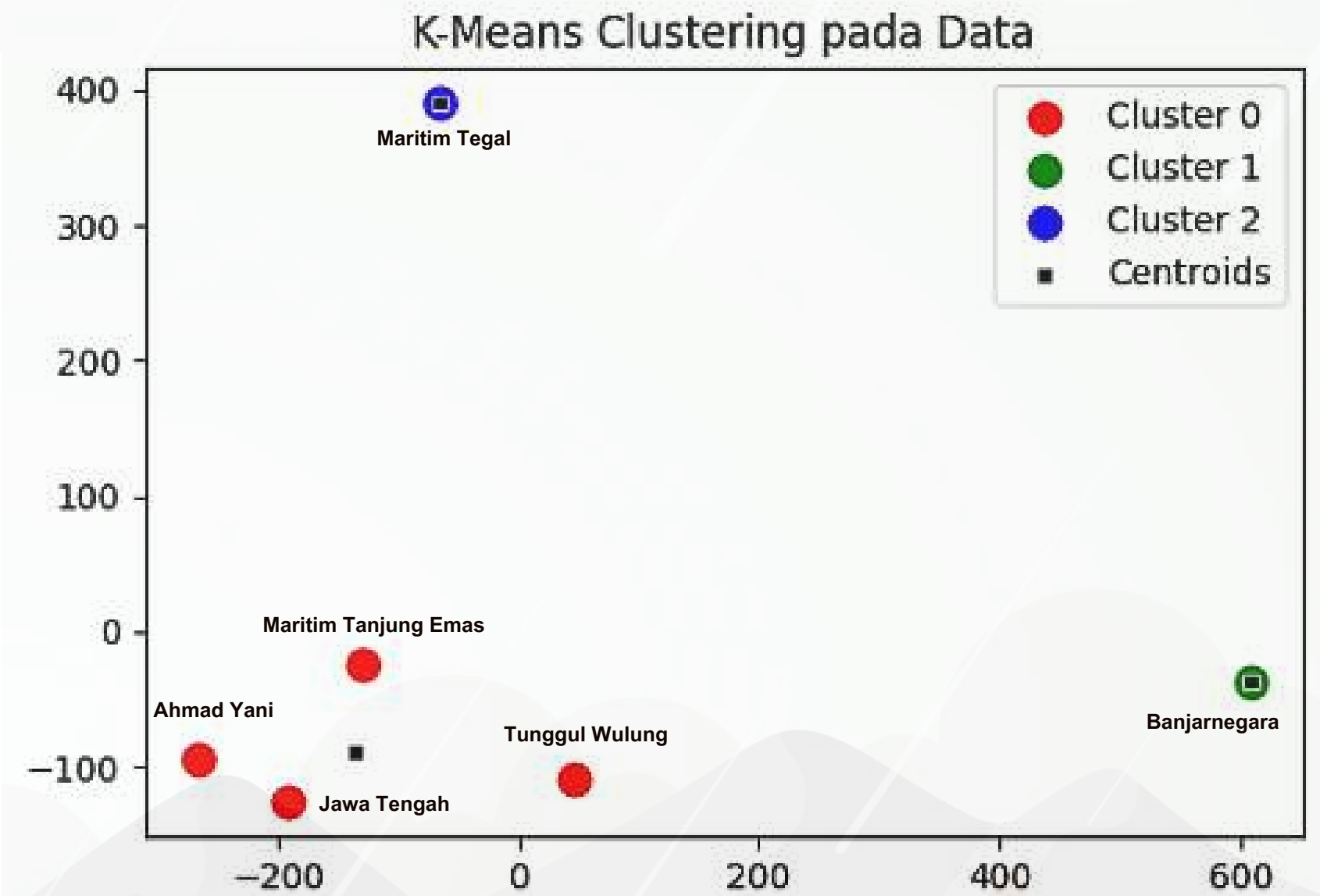
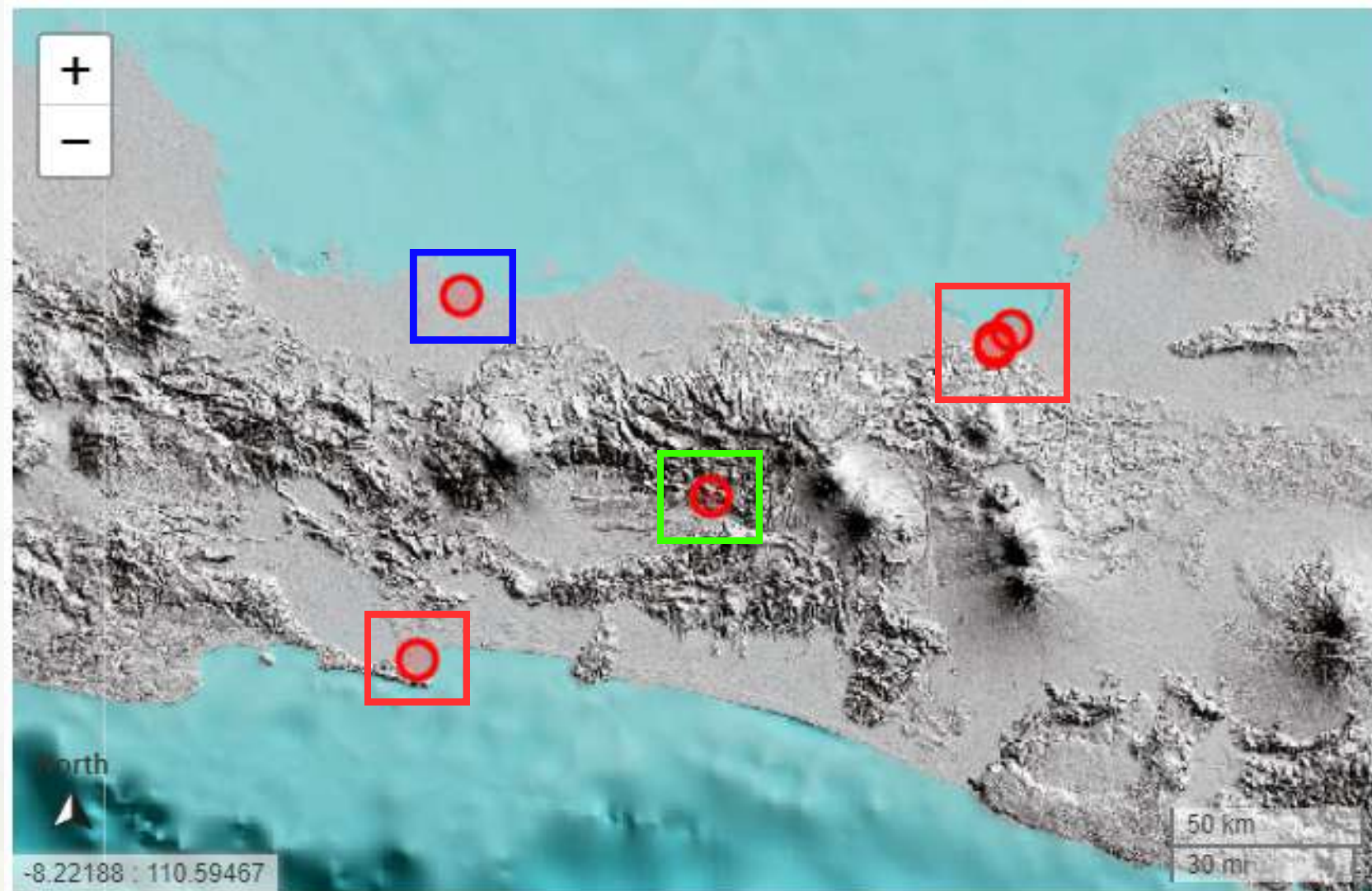
- Menetapkan cluster sebanyak tiga cluster.
- Memvisualisasikan tiap titik data dan cluster center pada scatter plot.
- Memberi warna tiap titik data berdasarkan masing-masing cluster.

| Stasiun | Sumbu x | Sumbu y | Cluster |
|-----------------------------|-------------|-------------|---------|
| Stamet Tunggul Wulung | 46.140764 | -109.451707 | 0 |
| Stageof Banjarnegara | 607.933335 | -36.223451 | 1 |
| Stamet Maritim Tegal | -65.123490 | 389.940428 | 2 |
| Staklim Jawa Tengah | -192.736440 | -125.908602 | 0 |
| Stamet Maritim Tanjung Emas | -129.297340 | -24.652412 | 0 |
| Stamet Ahmad Yani | -266.916829 | -93.704256 | 0 |

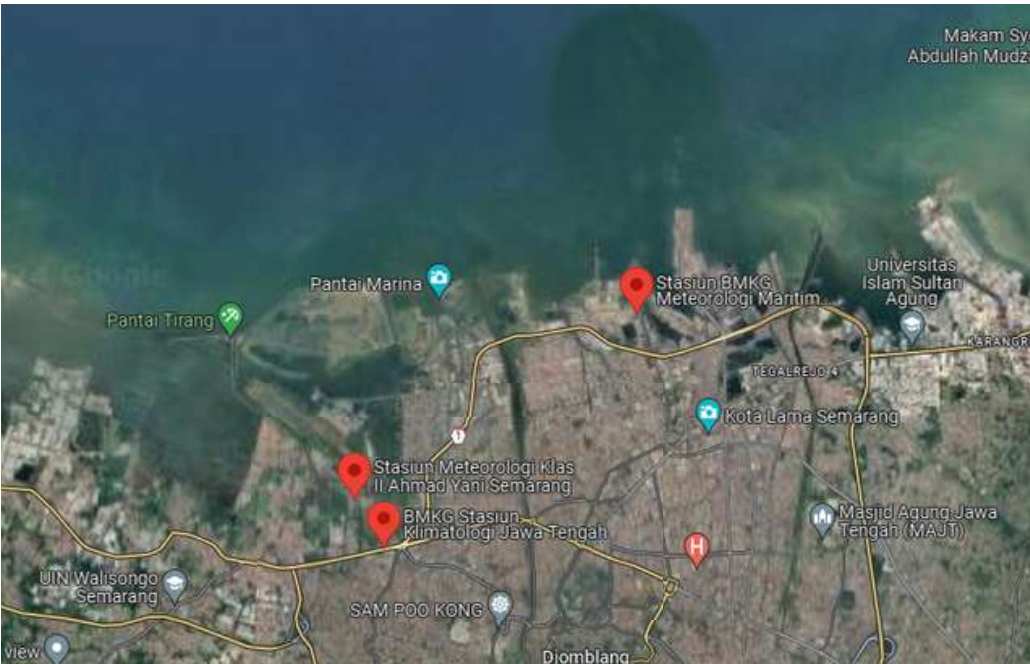
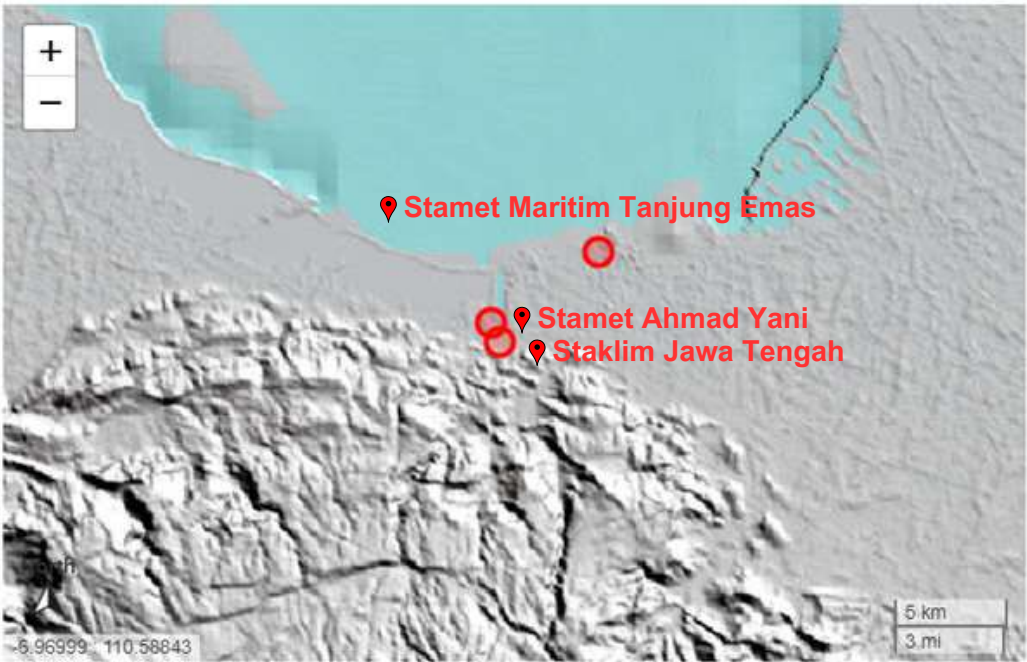


DETAIL TIAP CLUSTER

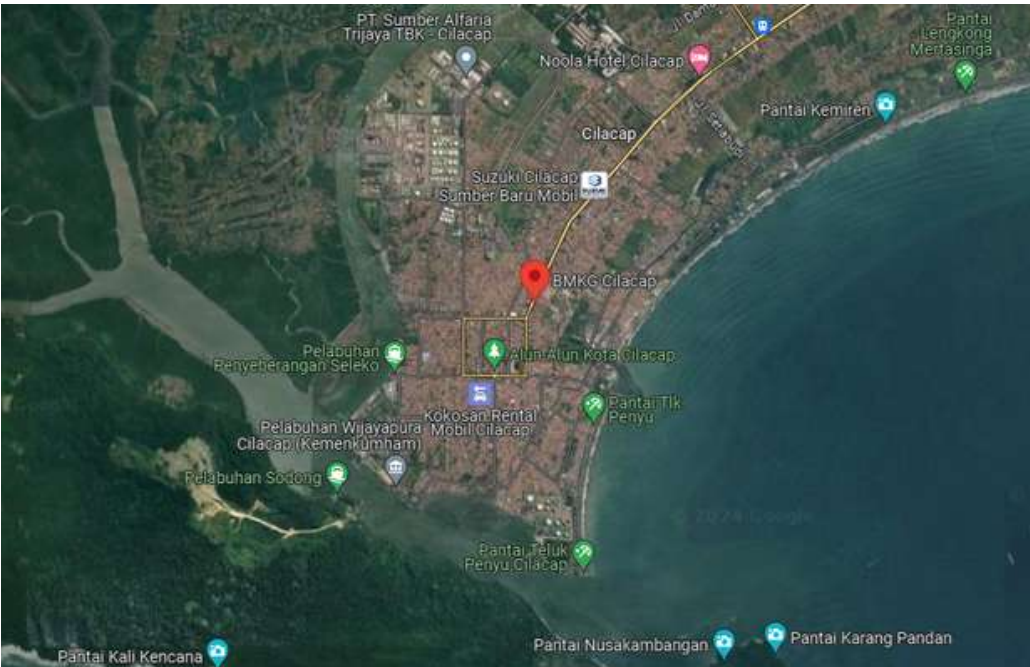
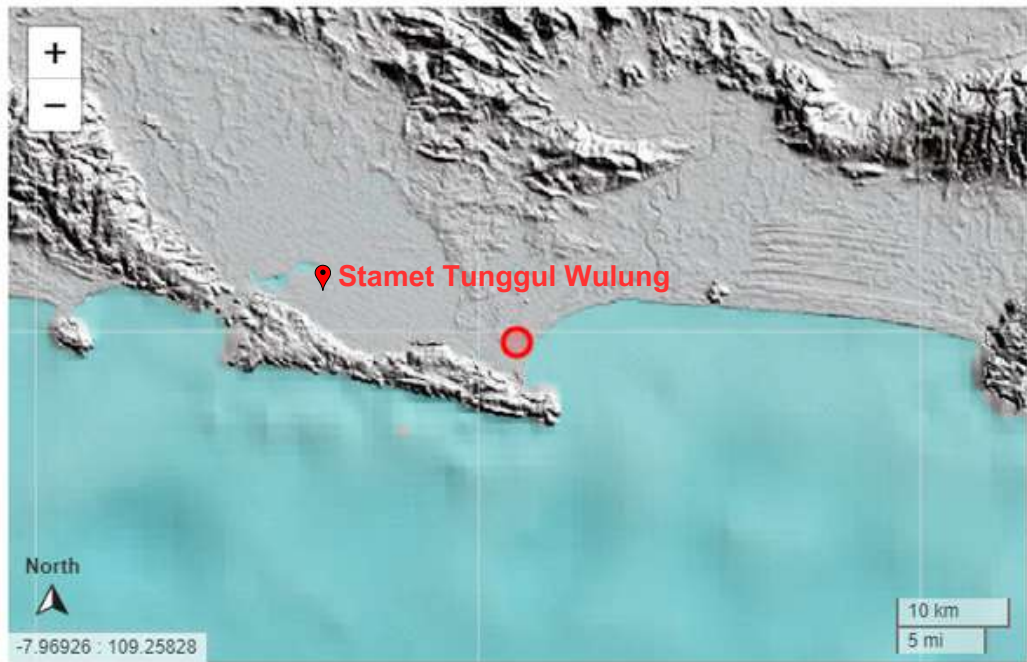
PETA PERSEBARAN STASIUN



PETA PERSEBARAN CLUSTER 0



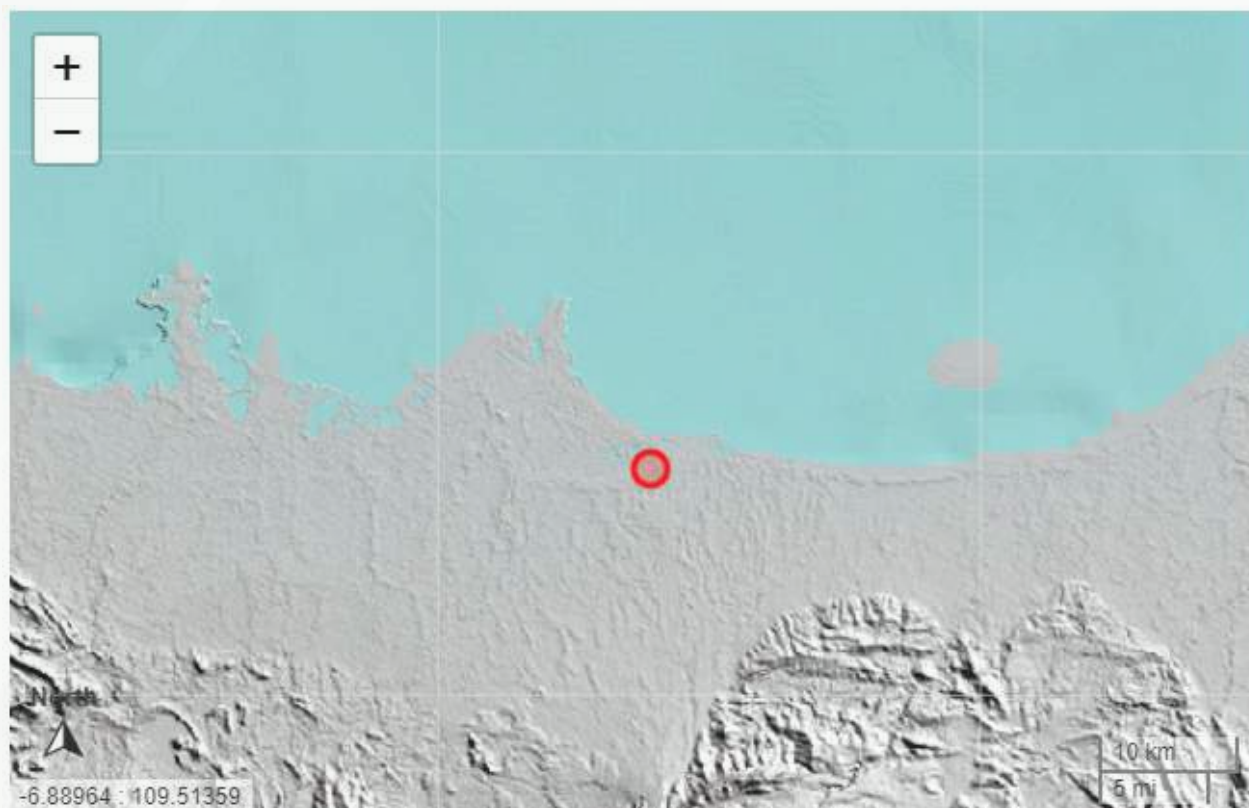
| Stasiun | Count | Mean | Std | Min | Max | Range |
|-----------------------------|-------|------|-----|-----|-----|-------|
| Stamet Tunggul Wulung | 10 | 289 | 103 | 139 | 444 | 305 |
| Staklim Jawa Tengah | 10 | 285 | 39 | 217 | 348 | 131 |
| Stamet Maritim Tanjung Emas | 10 | 280 | 56 | 206 | 370 | 164 |
| Stamet Ahmad Yani | 10 | 269 | 53 | 203 | 350 | 147 |



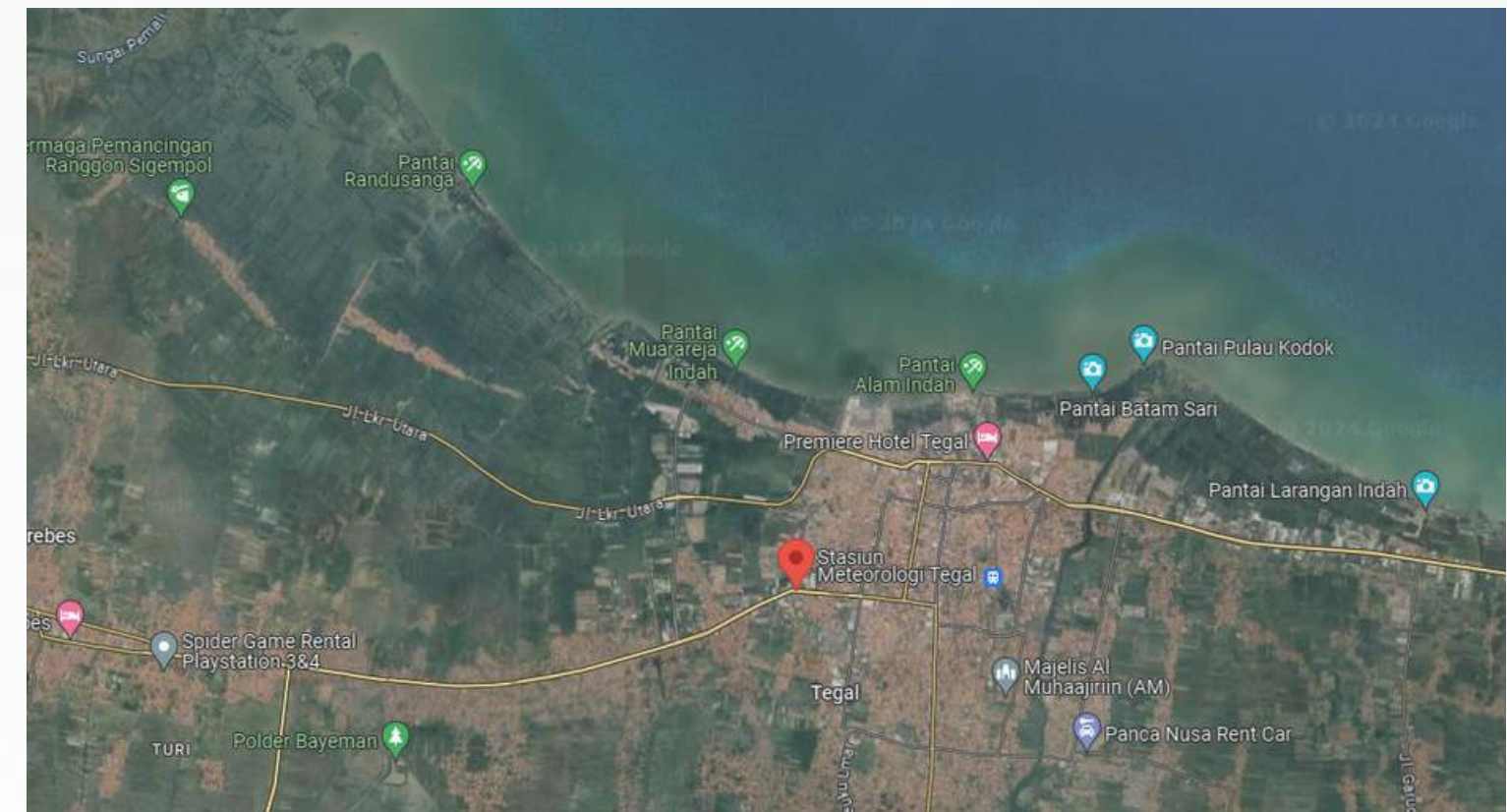
- Detail:
- Nilai rata-rata cenderung mirip.
 - Nilai standar deviasi mewakili sebaran curah hujan tiap tahun.
 - Jarak antara nilai minimum dan maksimum cenderung kecil.

PETA PERSEBARAN CLUSTER 2

📍 Stamet Maritim Tegal



| | |
|-------|-----|
| Count | 10 |
| Mean | 329 |
| Std | 146 |
| Min | 85 |
| Max | 551 |
| Range | 466 |

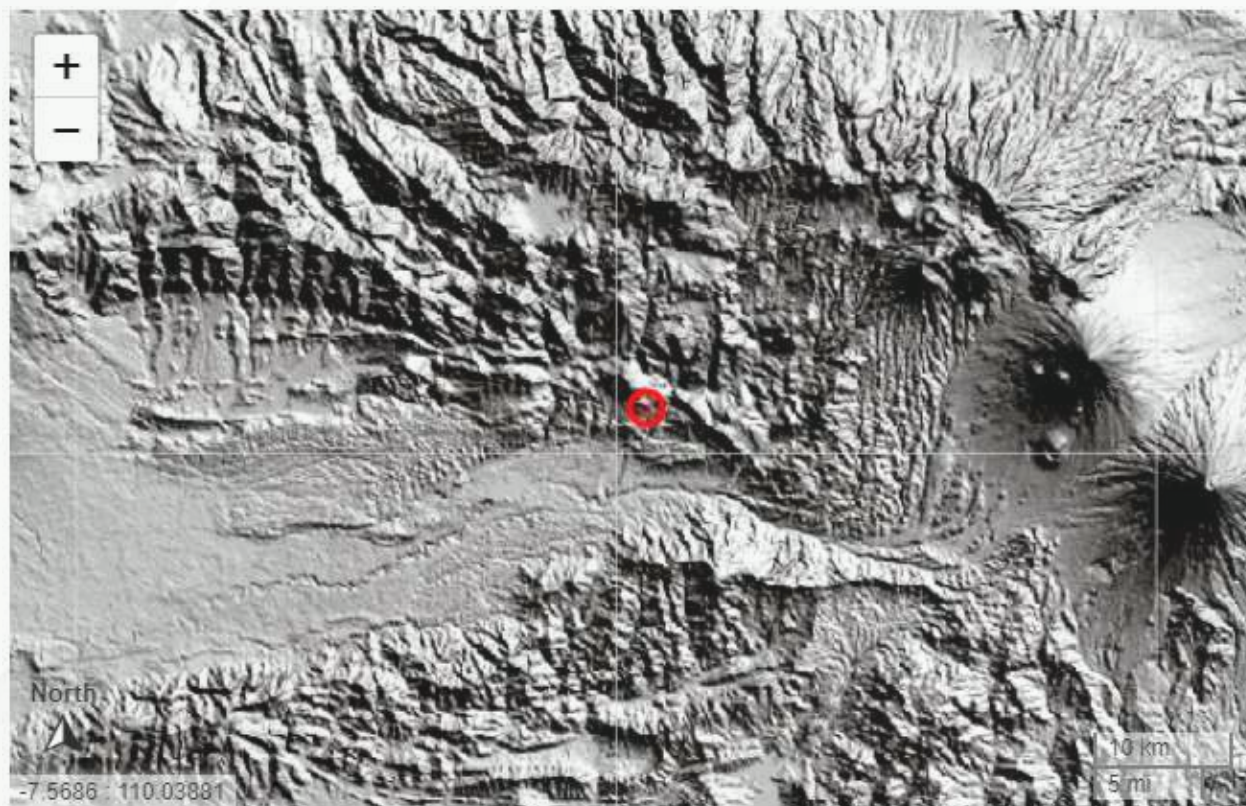


Detail:

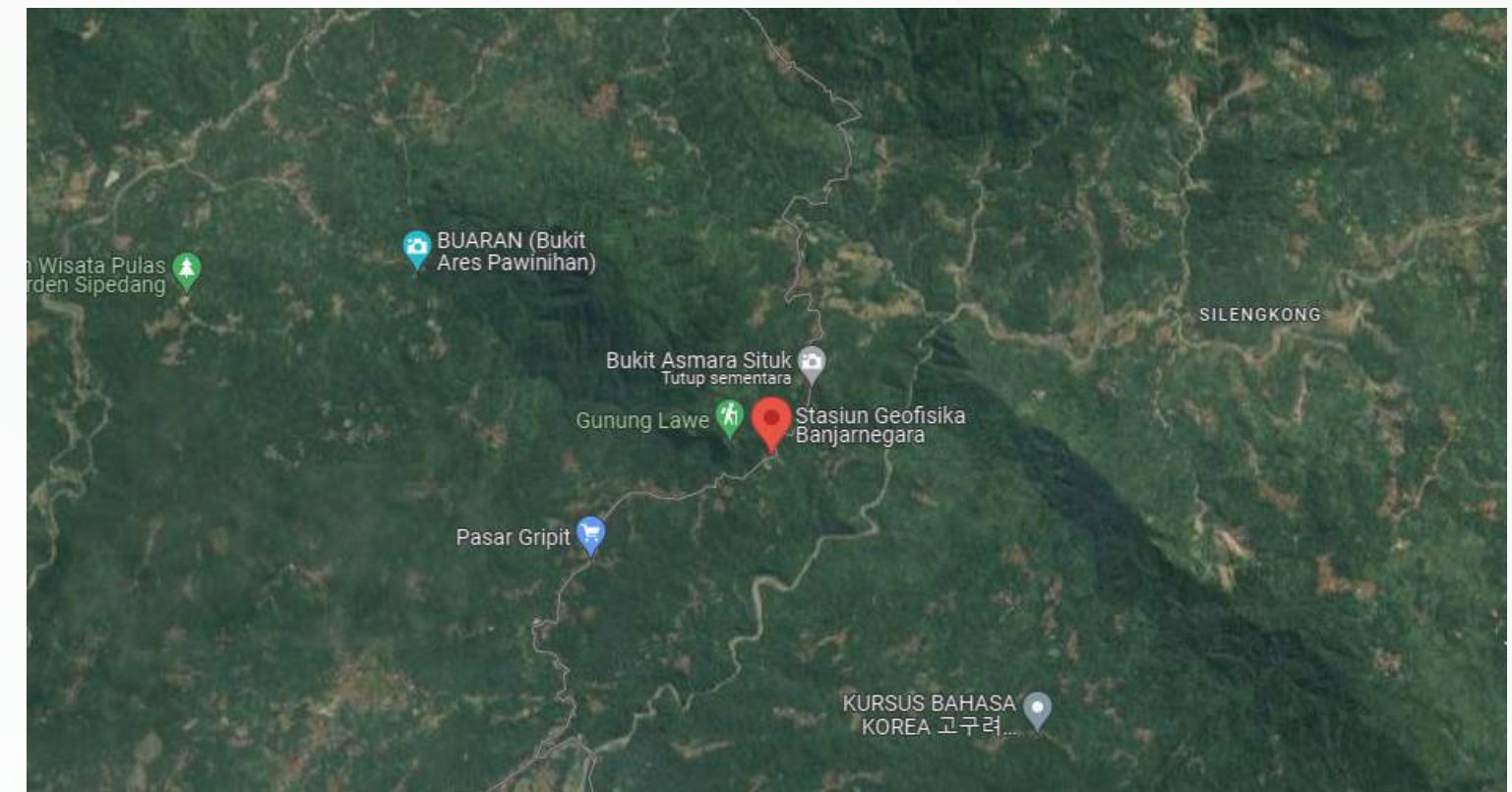
- Nilai rata-rata lebih besar dari cluster 0 dan lebih kecil dari cluster 1
- Nilai standar deviasi lebih besar dari cluster 0 dan lebih kecil dari cluster 1.
- Jarak antara nilai minimum dan maksimum cukup besar.

PETA PERSEBARAN CLUSTER 1

📍 Stageof Banjarnegara



| | |
|-------|-----|
| Count | 10 |
| Mean | 415 |
| Std | 213 |
| Min | 76 |
| Max | 765 |
| Range | 689 |



Detail:

- Nilai rata-rata lebih besar dari cluster 0 dan cluster 2.
- Nilai standar deviasi lebih besar dari cluster 0 dan cluster 2.
- Jarak antara nilai minimum dan maksimum besar.

KESIMPULAN

1. Dari hasil algoritma K-Means didapat **tiga jenis cluster optimal berdasarkan tingkat curah hujan** bulan Januari tahun 2015-2024 di Provinsi Jawa Tengah. Cluster 0 memiliki intensitas curah hujan yang lebih kecil dari Cluster 2 dan Cluster 1, Cluster 2 memiliki intensitas curah hujan yang lebih besar dari Cluster 0 dan lebih kecil dari Cluster 1, dan Cluster 1 memiliki intensitas curah hujan yang lebih tinggi dari Cluster 0 dan Cluster 2.
2. Tiap cluster memiliki perbedaan statistika deskriptif berupa **nilai rata-rata dan standar deviasi**. Cluster 0 memiliki nilai rata-rata dan standar deviasi curah hujan yang relatif kecil, Cluster 2 memiliki nilai rata-rata dan standar deviasi nilai curah hujan yang relatif sedang. Cluster 1 memiliki nilai rata-rata dan standar deviasi curah hujan yang relatif besar.
3. Perbedaan cluster curah hujan bulan Januari tahun 2015-2024 di Provinsi Jawa Tengah juga **dipengaruhi oleh kondisi geografis dan topografis** masing masing daerah.

DAFTAR PUSTAKA

Hasanah, Nur, Ugiarto, Muh. dan Puspitasar, Novianti. 2017. *Sistem Pengelompokan Curah Hujan Menggunakan Metode K-Means di Wilayah Kalimantan Timur*. Samarinda : Universitas Mulawarman

Izmi Asal, dan Pramono Hadi M.. 2016. *Efisiensi Jumlah Staisun Hujan untuk Analisis Hujan Tahubab di Provinsi Jawa Tengah dan Daerah Istimewa Yogyakarta*. Daerah Istimewa Yogyakarta : Core

Data Online Pusat Database - BMKG. Diakses pada 5 Februari 2024 dari https://dataonline.bmkg.go.id/data_iklim

**TERIMA
KASIH**

