



## Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Kluster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means

Putri Vania<sup>1</sup>, Betha Nurina Sari<sup>2</sup>

<sup>1,2</sup> Informatika, Universitas Singaperbangsa Karawang

### Abstract

Received: 23 Oktober 2023  
Revised : 28 Oktober 2023  
Accepted: 02 November 2023

Indonesia adalah negara agraris dengan industri pertanian yang besar, Indonesia memiliki hasil produksi padi tahun 2022 mencapai angka 54,75 juta ton GKG. Data yang dihasilkan dari sektor pertanian akan selalu bertambah banyak, maka perlu dilakukannya pengolahan data yang ada salah satunya yaitu menggunakan data mining. Proses data mining salah satunya adalah clustering, di mana dalam prosesnya perlu penentuan jumlah kluster yang tepat. Penelitian ini dilakukan dengan tujuan membandingkan metode untuk penentuan jumlah kluster yang optimal antara metode elbow dengan metode silhouette coefficient dalam pengelompokan produksi dan lahan padi di Jawa Barat pada tahun 2020, 2021, dan 2022 menggunakan algoritma k-means. Metodologi yang digunakan pada penelitian ini adalah Knowledge Discovery in Database (KDD). Kluster yang dihasilkan metode elbow berjumlah 3 kluster, sedangkan metode silhouette berjumlah 5 kluster. Hasil evaluasi DBI menggunakan silhouette coefficient sebesar 0,27 sedangkan metode elbow 0,39. Hasil penelitian ini menunjukkan metode silhouette coefficient lebih unggul dalam menentukan jumlah kluster dari pada metode elbow.

**Keywords:** Elbow, Clustering, K-Means, Perbandingan, Silhouette

(\*) Corresponding Author: [1910631170037@student.unsika.ac.id](mailto:1910631170037@student.unsika.ac.id)

**How to Cite:** Vania, P., & Sari, B. N. (2023). Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Kluster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means. <https://doi.org/10.5281/zenodo.10081332>

## PENDAHULUAN

Indonesia adalah negara agraris dengan industri pertanian yang besar dan hamparan lahan pertanian yang sangat luas. Padi merupakan salah satu tanaman pertanian terpenting di Indonesia. Menurut data BPS pada tahun 2022 produksi padi di Indonesia mencapai angka 54,75 juta ton GKG. Menurut Badan Pusat Statistik (BPS), Provinsi Jawa Barat merupakan salah satu penghasil padi terbesar, dengan total produksi padi pada tahun 2022 mencapai 9.433.723,09 Ton. Dengan pesatnya perkembangan pertanian, data yang dihasilkan akan semakin banyak seiring dengan berjalannya waktu. Data dengan jumlah yang banyak ini dapat dilakukan analisis salah satunya adalah menggunakan *data mining*. Pengolahan *data mining* ada berbagai macam diantaranya adalah asosiasi, klasifikasi, *clustering*, prediksi, dan regresi. Setiap data bisa diolah secara berbeda, seperti halnya *clustering* digunakan pada data yang bersifat numerik. *Clustering* digunakan untuk mengelompokkan data ke dalam beberapa kluster. Ada berbagai macam algoritma yang dapat digunakan pada proses *clustering* diantaranya yaitu algoritma *K-Means*, algoritma *K-Medoids*, dan algoritma *Fuzzy C-Means*. Setiap metode yang digunakan pada *clustering* tentunya memiliki kelebihan dan kekurangannya masing-masing.

*K-Means* adalah pendekatan pengelompokan di mana prosesnya diulang dan diulang sampai *centroid* tidak berubah. (Wahyudi, Sulthan, & Suhartini, 2021). *K-Means* dinilai mempunyai hasil yang lebih baik dibanding dengan algoritma yang lain yang biasanya digunakan pada *clustering*. Hal tersebut seperti pada penelitian yang telah dilakukan oleh Fitriana Harahap, 2021 di mana pada penelitiannya dilakukan perbandingan antara algoritma *k-means* dan algoritma *k-medoids* dalam *clustering* kelas siswa tunagrahita. Hasil dari penelitian yang dilakukan oleh Fitriana Harahap, 2021 yaitu algoritma *k-means clustering* dinilai memiliki hasil yang lebih unggul dengan nilai dbi 0,16 dibanding dengan algoritma *k-medoids*. Begitu pula pada penelitian yang dilakukan oleh Putri & Dwidayati, 2021 di mana dilakukan perbandingan algoritma *k-means* dan *fuzzy c-means* dalam mengelompokkan penyebaran covid-19. Hasil penelitian yang dilakukan oleh Putri & Dwidayati, 2021 adalah *K-Means* memiliki hasil akurasi yang lebih besar dibanding dengan *Fuzzy C-Means* dengan nilai dbi yang diperoleh *K-Means* sebesar 1,165219.

Mulai dari metode, karakteristik kumpulan data, struktur kepadatan data, ukuran data, dan jumlah kluster semuanya dapat memengaruhi hasil kluster yang (Dewi & Pramita, 2019). Untuk memilih jumlah kluster yang optimal pada *clustering* digunakan berbagai metode agar jumlah kluster yang dipilih memiliki hasil yang baik. Metode yang biasanya digunakan untuk memilih jumlah kluster yang tepat ada beberapa, seperti metode *elbow* dan juga metode *silhouette coefficient*. Metode *elbow* dan juga metode *silhouette coefficient* merupakan metode yang paling sering digunakan pada *clustering* untuk penentuan jumlah kluster yang tepat pada proses *data mining clustering*. *Elbow* maupun *silhouette coefficient* memiliki karakteristik masing-masing, di mana butuh ketepatan dalam memadukannya dengan metode yang digunakan pada proses *clustering* (Dewi & Pramita, 2019). Data kuantitatif yang diperoleh ialah data yang berasal dari Badan Pusat Statistik Jawa Barat. Informasi yang diterima berupa arsip data *excel* yang memuat jumlah padi yang diproduksi dan luas lahan padi di Jawa Barat pada tahun 2020, 2021, dan 2022.

Pada penelitian kali ini membandingkan antara metode *elbow* dan *silhouette coefficient* untuk penentuan jumlah kluster yang optimal pada *clustering*. Tahapan yang dilakukan pada penelitian ini mempergunakan *Knowledge Discovery in Database* (KDD). *Python* digunakan sebagai bahasa pemrograman dalam penelitian ini, dengan menggunakan Google Colab. Teknik yang digunakan adalah algoritma *k-means clustering*, yang dapat menghasilkan hasil *clustering* yang lebih baik berdasarkan penelitian sebelumnya.. Metode yang digunakan untuk mengevaluasi hasil *clustering* yang didapat yaitu menggunakan metode *Davies Bouldin Index* (DBI). Evaluasi dengan DBI dilakukan untuk melihat hasil perbandingan dari metode *elbow* dan juga metode *silhouette coefficient*.

## LANDASAN TEORI

### 1. Data Mining

*Data mining* ialah proses penggalian korelasi, pola, dan tren yang memiliki makna dari sekumpulan besar data (Jollyta, Ramdhan, & Zarlis, 2020). *Data mining* biasa digunakan untuk mengelola atau mencari pola pada sekumpulan data yang sangat besar.

## 2. Clustering

*Clustering* merupakan proses pengelompokan data berdasarkan keterkaitannya, dengan melihat kemiripan antar data dalam satu kelas dan perbedaan antar kelas/*cluster* (Jollyta, Ramdhan, & Zarlis, 2020).

## 3. KDD

Tahapan yang dilakukan pada penelitian ini menggunakan *Knowledge Discovery in Database* (KDD) yang memiliki 5 tahapan diantaranya *data selection* atau seleksi data, *preprocessing* data, transformasi data, *data mining*, dan evaluasi (Handoko, Fauziah, & Handayani, 2020).

## 4. K-Means

Dasar dari algoritma *k-means* adalah meminimalkan jarak atau biasa disebut *Sum of Squared Error* (SSE) antar objek yang ada pada data dengan mengikuti beberapa tahapan (Andini & Farokhah, 2022). Tahap-tahap tersebut adalah sebagai berikut:

1. Menentukan jumlah *k*.
2. Selanjutnya, Tentukanlah *centroid* secara acak.
3. Hitung jarak pada setiap data ke *centroid* dari setiap kluster *centroid* terdekat.
5. Lakukan Iterasi, kemudian tentukan posisi *centroid* baru.
6. Ulangi sampai *centroid* yang terbentuk tidak berubah lagi.

## 5. Elbow

Metode *elbow* ialah metode yang umumnya dipergunakan untuk penentuan jumlah kluster yang akan dibuat pada proses *clustering*. Metode *elbow* dipakai untuk memaksimalkan atau menentukan jumlah kluster yang optimal untuk proses *clustering* yang akan dilakukan. Cara kerja metode *elbow* adalah dengan menghitung kohesi dan pemisah kluster (Saputra, Saputra, & Oswari, 2019). Kohesi kluster mengukur seberapa erat keterkaitan data dalam klasternya sedangkan pemisah mengukur seberapa berbeda atau terpisahnya satu kluster dengan kluster lainnya. Kohesi kluster dan pemisahan kluster dapat diukur dengan menggunakan *sum of squares error* (SSE).

Dalam menghitung SSE dapat menggunakan rumus sebagai berikut :

$$SSE = \sum_{k=1}^k \sum_{x_i} |x_i - C_k|^2 \quad (1)$$

Keterangan :

- K = Kluster ke-C  
 $x_i$  = Jarak data pada obyek ke-i  
 $C_k$  = Pusat kluster ke-i

## 6. Silhouette Coefficient

Metode *Silhouette* menggunakan koefisien siluet yang menggabungkan pemisahan dan kohesi (Saputra, Saputra, & Oswari, 2019). Semakin besar koefisien *Silhouette*, semakin baik kluster tersebut.

$$SC = \max_k SI(k) \quad (1)$$

Keterangan :

$SC$  = *Silhouette Coefficient*

$SI$  = *Silhouette Index Global*

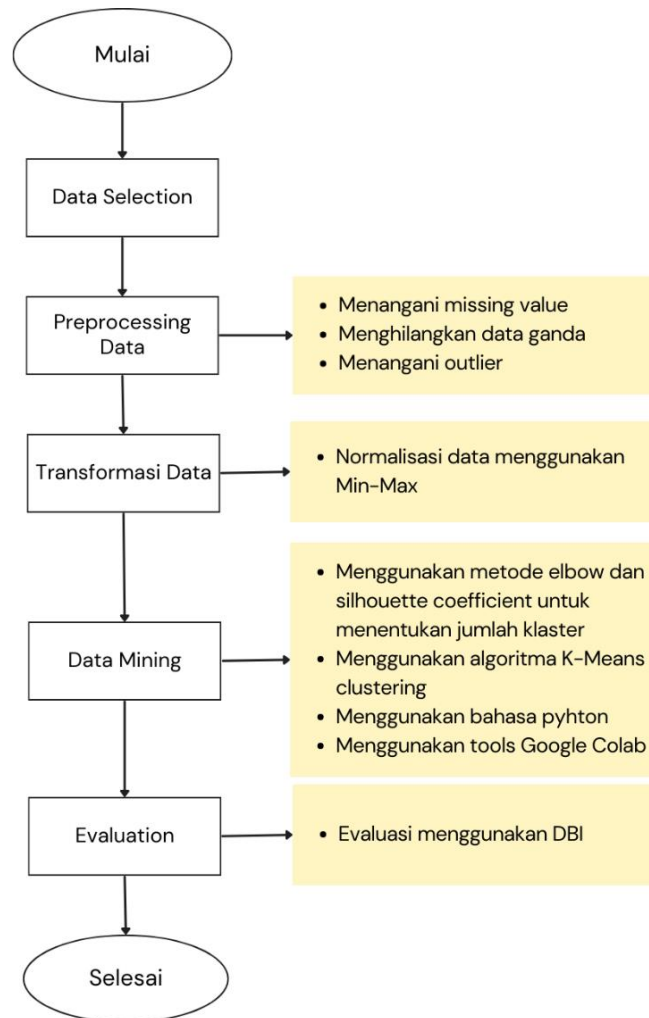
$k$  = jumlah klaster

## 7. DBI

DBI merupakan salah satu metode yang biasanya dipakai untuk mengevaluasi hasil clustering yang sudah dibuat. Penilaian Indeks *Davies-Bouldin* menampilkan sistem evaluasi klaster internal, di mana hasil klaster yang baik atau buruk ditentukan oleh jumlah dan kedekatan hasil klaster. Semakin mendekati 0 angka DBI nya semakin baik clustering yang dihasilkan. Pengukuran *Davies-Bouldin Indeks* ini berupaya untuk mengurangi jarak antar titik dalam suatu klaster sekaligus memaksimalkan jarak antar klaster. Jika jarak antar klaster paling besar, hal ini menunjukkan bahwa kemiripan atribut antar klaster sedang (Mughnyanti, Efendi, & Zarlis, 2020).

## METODE PENELITIAN

Metodologi yang dipergunakan pada penelitian kali ini ialah *Knowledge Discovery in Database* (KDD). Berikut merupakan rancangan dari penelitian ini :



**Gambar 1.** Rancangan Penelitian

### 1. Seleksi Data

Pada tahap seleksi data dilakukan pemilihan atribut yang relevan dengan penelitian. Hasil dari seleksi data ini yang selanjutnya akan diproses ke tahap *preprocessing data*. *Dataset* pada penelitian ini yaitu data luas dan produksi padi di Jawa Barat pada tahun 2020,2021, dan 2022.

### 2. *Preprocessing Data*

Setelah data melewati tahap seleksi, selanjutnya data akan masuk ke tahap *preprocessing* di mana data dilakukan pembersihan dari data ganda, nilai yang hilang, dan dilakukan penanganan *outlier* jika ditemukan pada data luas dan produksi padi di Jawa Barat. Data yang buruk akan memiliki hasil yang buruk, begitu pula dengan data yang baik akan memiliki hasil yang baik. Jadi tahap *preprocessing data* sangatlah penting.

### 3. Transformasi Data

Setelah data dibersihkan pada tahap *preprocessing*, selanjutnya data masuk ke dalam tahap transformasi data. Di mana jika data memiliki *range* berbeda akan dilakukan transformasi data agar memiliki *range* data yang sama. Normalisasi *Min-Max* digunakan untuk transformasi pada data luas dan produksi padi di Jawa Barat.

### 4. *Data Mining*

Setelah data dilakukan pemebersihan dan dilakukan transformasi, selanjutnya data masuk ke tahap inti yaitu *data mining*. Pada tahap ini dilakukan penentuan jumlah kluster yang akan dibuat mempergunakan metode *elbow* dan juga *silhouette coefficient*. Proses *clustering* pada data dengan menggunakan jumlah kluster dari masing-masing metode *elbow* dan *silhouette coefficient*. *Clustering* dilakukan dengan menggunakan algoritma *k-means* dan bahasa *python* pada google colab.

### 5. Evaluasi

Setelah *clustering* dilakukan akan dievaluasi hasil *clustering* dengan jumlah kluster *elbow* dan *silhouette coefficient* menggunakan DBI. Evaluasi dilakukan untuk bisa menilai dan membandingkan hasil dari jumlah kluster yang dipilih menggunakan metode *elbow* dan juga *silhouette coefficient*.

## HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 6 variabel yang merupakan jumlah produksi padi dan luas lahan padi di Jawa Barat pada tahun 2020,2021, 2022. Alat yang digunakan adalah google colab dengan bahasa pemrograman *python* untuk pengolahan datanya.

### Seleksi data

*Dataset* yang digunakan merupakan data jumlah produksi padi dan juga jumlah luas lahan pertanian padi di Jawa Barat pada tahun 2020,2021, dan 2022. Dilakukan pemilihan variabel apa saja yang akan digunakan untuk ke tahap *data mining*. Variabel dipilih berdasarkan keterkaitannya dengan topik. Seleksi data sangat penting karena akan memengaruhi hasil dari kluster yang akan dibuat.

**Tabel 1.** *Dataset* Penelitian

Kabupaten / Kota	Produksi 2020	Produksi 2021	Produksi 2022	Luas Lahan 2020	Luas Lahan 2021	Luas Lahan 2022
Bogor	298974.65	285154.00	299893.80	55692	57426	56002
Sukabumi	521459.25	492926.30	508220.48	93371	89510	95013
Cianjur	622992.32	611773.00	617941.03	113856	113539	115877
Bandung	277156.28	310715.20	289205.70	50150	50553	47538
Garut	424913.90	443319.90	441316.20	77873	75966	75951
Tasikmalaya	444358.24	445909.20	430961.73	72941	82935	81310
Ciamis	282791.46	320330.70	305676.17	52925	55013	55844
Kuningan	277317.90	275893.20	277583.76	45786	47236	48250
Cirebon	497080.10	464730.70	494699.98	83233	84946	84892
Majalengka	566334.48	534250.20	566087.34	98038	97516	100951
Sumedang	307178.00	262037.80	298163.00	55892	53341	56662
Indramayu	1363311.87	1319624.00	1482255.86	226626	227051	245329
Subang	970759.74	959456.40	1038780.58	169416	163947	177986
Purwakarta	159957.49	167865.30	188731.69	30316	32931	34397
Karawang	1087873.90	1234134.00	1226880.08	181915	197916	204326
Bekasi	504103.31	587586.10	555747.09	103172	100338	103089
Bandung Barat	164640.49	146494.60	153648.29	29197	27655	29202
Pangandaran	154982.57	140000.70	148670.61	29313	27678	30116
Kota Bogor	152.26	172.10	173.99	33	36	32
Kota Sukabumi	14532.67	15123.41	13254.34	2336	2456	2176
Kota Bandung	6799.08	7097.77	7186.11	1074	1026	995
Kota Cirebon	1031.24	1244.61	861.82	226	232	160
Kota Bekasi	2723.05	2304.54	1868.44	484	478	454
Kota Depok	189.23	244.35	72.13	31	45	15
Kota Cimahi	437.44	442.71	338.03	77	77	60
Kota Tasikmalaya	36349.80	49118.08	54233.66	7569	8540	9798
Kota Banjar	28371.86	35624.74	31271.18	5344	5723	5978

Pada Tabel 1 dapat dilihat di mana *dataset* yang telah dipilih dan akan digunakan pada tahap selanjutnya yaitu terdapat variabel kabupaten/kota, jumlah produksi dan luas lahan padi pada tahun 2020, 2021, 2022. Dengan jumlah 27 kabupaten/kota di provinsi Jawa Barat.

### Preprocessing data

Pada tahap *preprocessing* dilakukan penanganan *outlier*. Penanganan *outlier* menggunakan *interquartile range* dengan mengganti nilai *outlier* atas

dengan nilai batas atas dan nilai *outlier* bawah dengan nilai batas bawah. Penanganan *outlier* pada data tentunya dapat dilihat pada Gambar 2.

```

▶ outliers = df['produksi20']
  outliers1 = df['produksi21']
  outliers2 = df['produksi22']
  outliers3 = df['luas20']
  outliers4 = df['luas21']
  outliers5 = df['luas22']

# Hitung nilai Q1 dan Q3
q1, q3 = np.percentile(outliers, [25, 75])
q11, q31 = np.percentile(outliers1, [25, 75])
q12, q32 = np.percentile(outliers2, [25, 75])
q13, q33 = np.percentile(outliers3, [25, 75])
q14, q34 = np.percentile(outliers4, [25, 75])
q15, q35 = np.percentile(outliers5, [25, 75])

# Hitung nilai IQR
iqr = q3 - q1
iqr1 = q31 - q11
iqr2 = q32 - q12
iqr3 = q33 - q13
iqr4 = q34 - q14
iqr5 = q35 - q15
    
```

**Gambar 2.** Penanganan *Outlier*

### Transformasi Data

Data yang sudah bersih dan sudah melalui tahap *preprocessing* selanjutnya masuk ke tahap transformasi data di mana data yang ada akan dilakukan transformasi agar range data yang satu dengan yang lainnya tidaklah berbeda.

```

▶ df['produksi20'] = (df['produksi20']-df['produksi20'].min())/(df['produksi20'].max()-df['produksi20'].min())
  df['produksi21'] = (df['produksi21']-df['produksi21'].min())/(df['produksi21'].max()-df['produksi21'].min())
  df['produksi22'] = (df['produksi22']-df['produksi22'].min())/(df['produksi22'].max()-df['produksi22'].min())
  df['luas20'] = (df['luas20']-df['luas20'].min())/(df['luas20'].max()-df['luas20'].min())
  df['luas21'] = (df['luas21']-df['luas21'].min())/(df['luas21'].max()-df['luas21'].min())
  df['luas22'] = (df['luas22']-df['luas22'].min())/(df['luas22'].max()-df['luas22'].min())

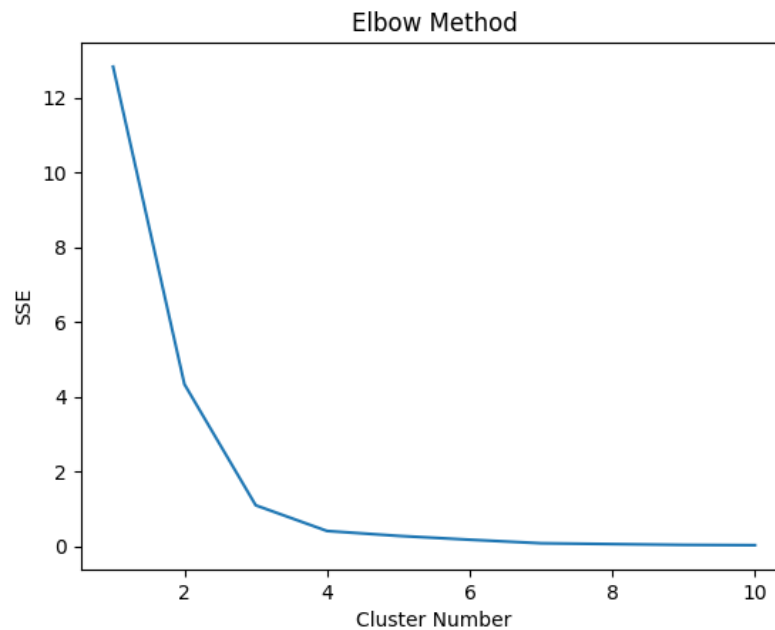
df
    
```

**Gambar 3.** *Min-Max Normalization*

Seperti yang tertera pada Gambar 3, dilakukan transformasi data menggunakan *min-max normalization*. *Min-Max normalization* akan mengubah range data menjadi 0-1.

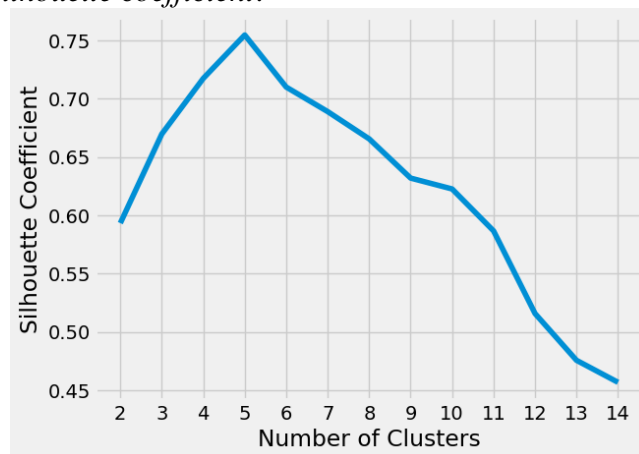
### Data Mining

Data yang sudah siap akan diproses pada langkah penambangan data. Pertama, tentukan berapa banyak kluster yang akan diproduksi. Pendekatan *elbow* dan metode *silhouette coefficient* digunakan untuk penentuan jumlah kluster.



**Gambar 4.** Hasil Metode *Elbow*

Seperti yang ada pada Gambar 4, hasil dari metode *elbow* menunjukkan bahwa jumlah kluster yang optimal pada *clustering* jumlah produksi padi dan luas lahan padi di Jawa Barat pada tahun 2020,2021, dan 2022 adalah dengan dibuat menjadi 3 kluster. Selanjutnya dilakukan juga penentuan jumlah kluster menggunakan *silhouette coefficient*.



**Gambar 5.** Hasil dari Silhouette Coefficient

Pada Gambar 5 terlihat bahwa jumlah kluster yang optimal berdasarkan metode *silhouette coefficient* adalah 5 kluster. Selanjutnya akan dilakukan pengclustering data dengan menggunakan jumlah kluster yang sudah didapat tadi dari hasil metode *elbow* dan juga metode *silhouette coefficient*.

	kabupaten	cluster
0	Bogor	1
1	Sukabumi	1
2	Cianjur	1
3	Bandung	1
4	Garut	1
5	Tasikmalaya	1
6	Ciamis	1
7	Kuningan	1
8	Cirebon	1
9	Majalengka	1
10	Sumedang	1
11	Indramayu	2
12	Subang	2
13	Purwakarta	3
14	Karawang	2
15	Bekasi	1
16	Bandung Barat	3
17	Pangandaran	3
18	Kota Bogor	3
19	Kota Sukabumi	3
20	Kota Bandung	3
21	Kota Cirebon	3
22	Kota Bekasi	3
23	Kota Depok	3
24	Kota Cimahi	3
25	Kota Tasikmalaya	3
26	Kota Banjar	3

**Gambar 6.** Hasil *Clustering* dengan 3 klaster

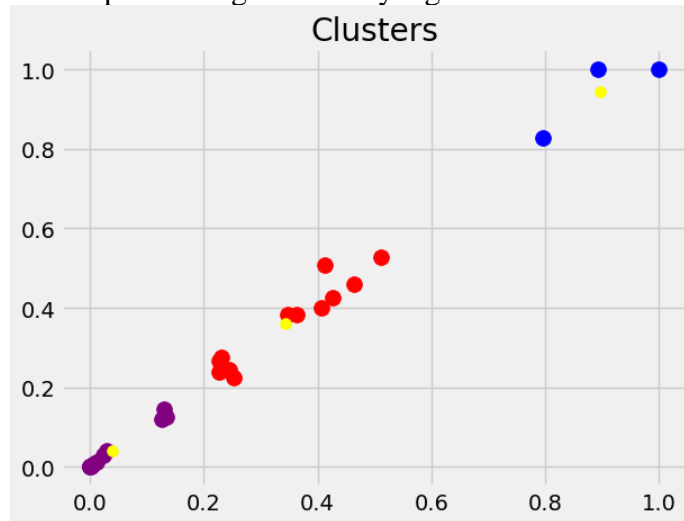
Pada gambar 6, merupakan hasil *clustering* dengan jumlah 3 klaster menggunakan algoritma *k-means* berdasarkan metode *elbow* yang sudah dilakukan. Didapatkan hasil 12 Kabupaten di klaster 1, 3 kabupaten di klaster 2, dan 12 Kabupaten/kota di klaster 3. Selanjutnya akan dilakukan *clustering* dengan jumlah klaster yang didapat dari *silhouette coefficient* yaitu berjumlah 5 klaster.

	kabupaten	cluster
0	Bogor	1
1	Sukabumi	4
2	Cianjur	4
3	Bandung	1
4	Garut	4
5	Tasikmalaya	4
6	Ciamis	1
7	Kuningan	1
8	Cirebon	4
9	Majalengka	4
10	Sumedang	1
11	Indramayu	2
12	Subang	2
13	Purwakarta	5
14	Karawang	2
15	Bekasi	4
16	Bandung Barat	5
17	Pangandaran	5
18	Kota Bogor	3
19	Kota Sukabumi	3
20	Kota Bandung	3
21	Kota Cirebon	3
22	Kota Bekasi	3
23	Kota Depok	3
24	Kota Cimahi	3
25	Kota Tasikmalaya	3
26	Kota Banjar	3

**Gambar 7.** Hasil *Clustering* dengan 5 Klaster

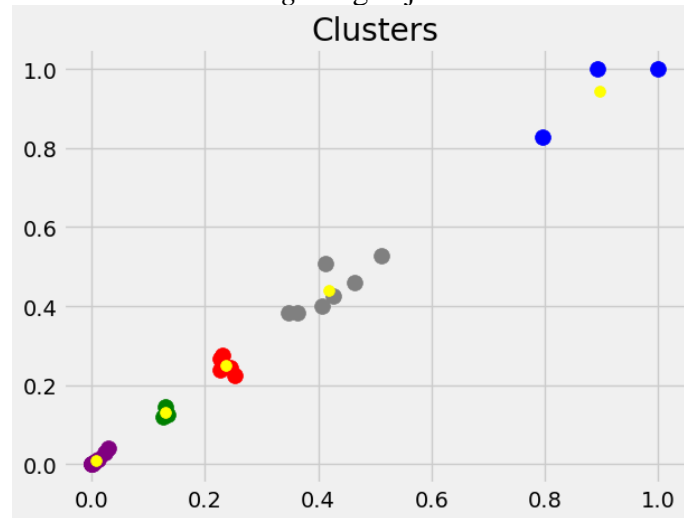
Pada Gambar 7 terdapat hasil dari *clustering* dengan jumlah 5 klaster menggunakan algoritma *k-means*, jumlah klaster ini berdasarkan *silhouette*

*coefficient* yang sudah dilakukan sebelumnya. Terdapat 5 kabupaten/kota di klaster 1, 3 kabupaten/kota di klaster 2, 9 kabupaten/kota di klaster 3, 7 kabupaten/kota di klaster 4, dan 3 kabupaten/kota di klaster 5. Selanjutnya dilakukan visualisasi hasil klaster untuk melihat perbandingan klaster yang telah terbentuk.



**Gambar 8.** Visualisasi dengan Jumlah 3 Klaster

Pada gambar 8 merupakan hasil visualisasi *clustering* dengan jumlah 3 klaster berdasarkan metode *elbow* yang sudah dilakukan sebelumnya. Selanjutnya dilakukan visualisasi hasil *clustering* dengan jumlah 5 klaster untuk perbandingan.



**Gambar 9.** Visualisasi dengan Jumlah 5 Klaster

Pada gambar 9 dapat dilihat hasil visualisasi *clustering* dengan jumlah 5 klaster berdasarkan metode *silhouette coefficient* yang sudah dilakukan. Jika dibandingkan dengan visualisasi dengan jumlah 3 klaster, klaster terbagi lebih baik saat dibagi menjadi 5 klaster. Tetapi belum bisa dipastikan jika hanya melihat dari visualisasi tadi, maka dilakukanlah evaluasi menggunakan DBI untuk mengetahui apakah *clustering* sudah terpisah dengan baik dan juga untuk membandingkan hasil klaster dengan jumlah 3 klaster dan 5 klaster.

```

▶ from sklearn.metrics import davies_bouldin_score
  db_index = davies_bouldin_score(x, y_kmeans)
  print(db_index)

```

0.39899258942107635

**Gambar 10.** Hasil DBI dengan Jumlah 3 Klaster

```

▶ from sklearn.metrics import davies_bouldin_score
  db_index = davies_bouldin_score(x, y_kmeans)
  print(db_index)

```

0.2700874848272372

**Gambar 11.** Hasil DBI dengan Jumlah 5 Klaster

Seperti yang ada pada gambar 10 dan 11, di mana nilai DBI dengan jumlah 3 klaster adalah 0,3989 dan nilai DBI dengan jumlah 5 klaster adalah 0,27. Nilai tersebut menunjukkan bahwa data produksi padi dan luas lahan padi pada tahun 2020,2021,2022 memiliki hasil yang lebih baik jika dibagi menjadi 5 klaster dengan perolehan nilai DBI lebih mendekati 0 yaitu 0.27. Jadi, dapat disimpulkan bahwa metode *silhouette coefficient* lebih unggul dalam menentukan jumlah klaster yang optimal dibandingkan dengan metode *elbow* karena hasil DBI dengan jumlah 5 klaster yang diperoleh dari metode *silhouette coefficient* lebih mendekati 0.

## KESIMPULAN

Berdasarkan hasil yang diperoleh dari serangkaian penelitian yang telah dilakukan dengan membandingkan metode *elbow* dan juga metode *silhouette coefficient* dengan menggunakan algoritma *k-means* dalam mengelompokkan jumlah produksi padi dan luas lahan padi di Jawa Barat pada tahun 2020,2021,2022. Pada penelitian ini dapat disimpulkan bahwa metode *silhouette coefficient* lebih unggul dibandingkan dengan *elbow* dalam menentukan jumlah klaster yang optimal pada *clustering* jumlah produksi padi dan luas lahan padi di Jawa Barat pada tahun 2020,2021, dan 2022 menggunakan algoritma *k-means*. Hal tersebut dibuktikan dengan nilai DBI yang diperoleh jumlah klaster *silhouette coefficient* lebih rendah dengan nilai 0,27 dibandingkan dengan jumlah klaster *elbow* yaitu 0,39.

## DAFTAR PUSTAKA

- Andini, T. D., & Farokhah, L. (2022). Peningkatan Ketersediaan darah sesuai segmentasi Umur Menggunakan K-means clustering. *Jurnal Manajemen Informatika (JAMIKA)*, 12(2), 126–136. <https://doi.org/10.34010/jamika.v12i2.7897>
- Dewi, D. A., & Pramita, D. A. (2019). Analisis Perbandingan metode elbow Dan Silhouette Pada algoritma clustering K-medoids Dalam Pengelompokan Produksi Kerajinan bali. *Matrix : Jurnal Manajemen Teknologi Dan Informatika*, 9(3), 102–109. <https://doi.org/10.31940/matrix.v9i3.1662>
- Handoko, S., Fauziah, F., & Handayani, E. T. (2020). Implementasi data mining untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-means clustering. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>

- Harahap, F. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Clustering Kelas Siswa Tunagrahita. *TIN: Terapan Informatika Nusantara*, 2(4), 191-197.
- Jollyta, D., Ramdhan, W., & Zarlis, M. Konsep Data Mining Dan Penerapan, Sleman: Deepublish, 2020.
- Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of determining centroid clustering x-means algorithm with Davies-Bouldin Index Evaluation. *IOP Conference Series: Materials Science and Engineering*, 725(1), 012128. <https://doi.org/10.1088/1757-899x/725/1/012128>
- Saputra, D. M., Saputra, D., & Oswari, L. D. (2020). Effect of distance metrics in determining K-value in k-means clustering using elbow and silhouette method. *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*. <https://doi.org/10.2991/aisr.k.200424.051>
- Wahyudi, I., Sulthan, M. B., & Suhartini, L. (2021). Analisa Penentuan cluster Terbaik Pada metode k-means menggunakan elbow terhadap sentra Industri Produksi di Pamekasan. *Jurnal Aplikasi Teknologi Informasi Dan Manajemen (JATIM)*, 2(2), 72–81. <https://doi.org/10.31102/jatim.v2i2.1274>