

# ANALISIS CLUSTERING PENYEBARAN HIV DI KARAWANG BERDASARKAN KECAMATAN DENGAN ALGORITMA K-MEANS

Dica Parameswari Syifa Dewi<sup>1\*</sup>, Bagja Nugraha<sup>2</sup>, Siska<sup>3</sup>

<sup>1,2,3</sup>Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361; Telp.(0267) 641177

Received: 11 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

## Keywords:

HIV; K-MEANS; KDD; QGIS; Silhouette Coefficient

## Correspondent Email:

20106131250008@student.unsika.ac.id

**Abstrak.** Pergaulan bebas semakin marak seiring perkembangan zaman. Perilaku ini melibatkan hubungan seksual yang berdampak negatif pada individu dan masyarakat. Contohnya, peningkatan kasus penyakit menular seksual seperti HIV yang membahayakan kesehatan. HIV, singkatan dari Human Immunodeficiency Virus, melemahkan sistem kekebalan tubuh. Penelitian ini bertujuan mengetahui kluster kecamatan dengan penyebaran HIV tertinggi di Kabupaten Karawang menggunakan algoritma K-Means dan metodologi KDD. Dengan bantuan dari *tools Jupyter Notebook* maka hasil dari penelitian ini adalah terbentuknya  $k=3$ , yaitu  $k=0$  rawan,  $k=1$  tidak rawan, dan  $k=2$  sangat rawan, yang akan diolah dengan Algoritma *K-Means*. Hasil dari penerapan algoritma *K-Means* adalah dari 77 puskesmas dan rumah sakit, 42 berada di kluster 0, 24 di kluster 1, dan 11 di kluster 2. Hasil dari *clustering* dievaluasi dengan *Silhouette Coefficient* dan menghasilkan nilai sebesar 0.634 hasil tersebut menunjukkan bahwa kluster berstruktur baik. Setelah selesai melakukan evaluasi, hasil tersebut akan dilakukan visualisasi pemetaan menggunakan *tools Qgis*.

**Abstract.** Promiscuity is becoming more prevalent with the times. This behavior involves sexual relations that have a negative impact on individuals and society. For example, there is an increase in cases of sexually transmitted diseases such as HIV that endanger health. HIV, which stands for Human Immunodeficiency Virus, weakens the immune system. This study aims to determine the sub-district cluster with the highest HIV spread in Karawang Regency using the K-Means algorithm and KDD methodology. With the help of *Jupyter Notebook* tools, the result of this research is the formation of  $k = 3$ , namely  $k = 0$  vulnerable,  $k = 1$  not vulnerable, and  $k = 2$  very vulnerable, which will be processed with the K-Means Algorithm. The results of the application of the K-Means algorithm are from 77 health centers and hospitals, 42 are in cluster 0, 24 in cluster 1, and 11 in cluster 2. The results of clustering are evaluated with *Silhouette Coefficient* and produce a value of 0.634, these results indicate that the cluster is well structured. After completing the evaluation, the results will be visualized using *Qgis* tools.

## 1. PENDAHULUAN

Semakin berkembangnya zaman, perilaku pergaulan bebas semakin marak di kalangan umum. Pergaulan bebas adalah perilaku sosial di mana individu melakukan hubungan seksual tanpa mempertimbangkan nilai moral dan

norma yang berlaku dalam masyarakat. Fenomena ini sering dihubungkan dengan gaya hidup yang liberal dan kebebasan individu dalam mengambil keputusan tentang seksualitasnya.

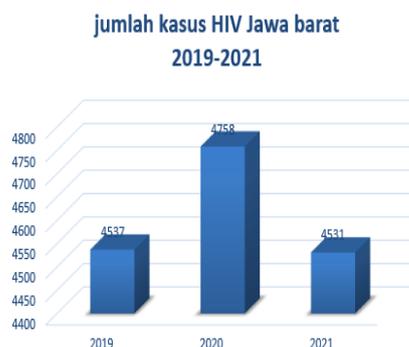
Namun, pergaulan bebas juga memiliki dampak yang merugikan bagi individu dan masyarakat. Salah satu contoh dari pergaulan bebas adalah semakin meningkatnya kasus penyakit menular seksual seperti HIV, yang dapat membahayakan kesehatan dan kehidupan individu.

HIV atau kepanjangan dari Human Immunodeficiency Virus adalah virus yang melemahkan sistem pertahanan tubuh terhadap penyakit. Terlepas dari fakta bahwa tidak ada obat HIV yang jelas, terdapat metode yang dapat digunakan untuk mengurangi perkembangan penyakit [1].

Virus mengganggu sistem kekebalan tubuh dengan merusak sel darah putih, terutama limfosit. Terdapat beberapa orang bertahan hidup tanpa gejala seperti seorang penderita HIV, bahkan ada beberapa penderita HIV yang tidak memerlukan pengobatan. HIV tidak dapat menyebar melalui cairan tubuh seperti air liur atau air kencing tetapi HIV dapat ditularkan melalui beberapa faktor, yaitu hubungan seksual yang memiliki risiko dan berbagi jarum suntik bekas dengan orang lain [2].

Dalam 3 tahun kebelakang menurut data dari Dinas Kesehatan Jawa Barat, yaitu pada tahun 2019-2021, terdapat 19.101 orang terinfeksi HIV di Provinsi Jawa Barat. Jumlah infeksi sebanyak itu menunjukkan adanya masalah serius terkait dengan penyebaran HIV di wilayah tersebut. Hal ini memerlukan perhatian khusus dari pemerintah dan masyarakat untuk melakukan tindakan guna mengendalikan penyebaran penyakit ini. Selain itu, perlu juga dilakukan sosialisasi dan edukasi kepada masyarakat mengenai bahaya HIV.

Untuk rincian dari kasus HIV setiap tahunnya, terdapat pada gambar dibawah ini:



**Gambar 1.** Jumlah Kasus HIV di Jawa Barat 2019-2021

Berdasarkan dengan latar belakang diatas, maka penelitian bertujuan untuk menentukan daerah mana yang memiliki penyebaran kasus HIV yang tertinggi di Kabupaten Karawang. Diharapkan dengan adanya penelitian ini, akan memberikan masukan untuk pemerintah Kabupaten Karawang mengenai perhatian yang lebih serius terhadap daerah-daerah yang menjadi klaster tertinggi.

Penelitian sebelumnya yang dilakukan oleh [3] membahas penggunaan algoritma K-Means dalam mengelompokkan jumlah kasus penyakit AIDS berdasarkan provinsi. Penelitian ini bertujuan untuk memberikan wawasan kepada pemerintah terkait provinsi-provinsi dengan jumlah kasus penyakit AIDS yang tinggi. Hasil penelitian menunjukkan bahwa beberapa provinsi di Indonesia, seperti Papua, DKI Jakarta, Jawa Barat, dan Jawa Timur, memiliki jumlah kasus penyakit AIDS yang signifikan. Metode K-Means digunakan untuk mengelompokkan data kasus penyakit AIDS berdasarkan provinsi, dengan harapan dapat memberikan masukan kepada pemerintah untuk memberikan perhatian lebih terhadap provinsi-provinsi dengan jumlah kasus penyakit AIDS yang tinggi. Penelitian ini memiliki potensi untuk memberikan kontribusi pada upaya penanggulangan penyebaran penyakit AIDS di Indonesia.

Penelitian lainnya yang dilakukan oleh [4], [5], [6], [7], [8] dalam jurnal-jurnal mereka menggunakan metode analisis cluster, khususnya K-Means atau Fuzzy C-Means, untuk mengelompokkan entitas geografis atau individu berdasarkan tingkat prevalensi atau kasus penyakit menular, terutama HIV/AIDS. Mereka memanfaatkan data sekunder, seperti Laporan Situasi HIV, PIMS Dinas Kesehatan Provinsi, dan laporan tahunan Kementerian Kesehatan RI, untuk tahun-tahun tertentu (2017, 2022, 2019, dan 2020). Hasil penelitian menunjukkan adanya 3 cluster optimal dengan variasi tingkat kasus, seperti kabupaten/kota dalam kategori siaga, standby, atau awas pada tingkat level epidemik HIV atau kasus penyakit menular lainnya. Rekomendasi yang diberikan melibatkan prioritas skrining/tes, konseling, serta penyediaan layanan perawatan dan dukungan bagi individu yang hidup dengan HIV. Penelitian ini memberikan kontribusi dalam merencanakan program pencegahan, pengendalian penyakit, dan pemahaman

distribusi serta perilaku populasi yang terinfeksi, dengan tujuan mengembangkan strategi efektif untuk mengatasi penyakit menular di tingkat geografis atau individu.

Penelitian ini disusun ke dalam beberapa kelompok klaster menggunakan metode K-means Clustering. Untuk menentukan klaster terbaik akan digunakan metode Elbow dengan menghitung nilai dari Sum of Square Error (SSE) dari masing-masing klaster. Proses evaluasi dari hasil klaster yang didapatkan akan menggunakan evaluasi Silhouette Coefficient untuk mengevaluasi seberapa baik data dikelompokkan dalam satu klaster.

## 2. TINJAUAN PUSTAKA

### 2.1 K-Means Clustering

Klastering adalah suatu metode atau teknik yang memanfaatkan pengelompokan data sesuai dengan kategori ataupun kelasnya masing-masing, sehingga data yang memiliki kesamaan akan dikelompokkan bersama dalam satu kelas data [9].

K-Means adalah salah satu algoritma yang paling banyak diterapkan dalam machine learning dan data mining. Algoritma K-Means bertujuan untuk mengelompokkan data ke dalam beberapa kelompok (cluster) yang mempunyai perbedaan antara satu dengan yang lain, berdasarkan fitur atau ciri-ciri yang dipunyai oleh setiap data [10].

### 2.2 Silhouette Coefficient

Untuk menilai kekuatan dan kualitas klaster yang dihasilkan, penting untuk memvalidasi hasil klaster tersebut. Salah satu cara untuk melakukan validasi adalah menggunakan *Silhouette Coefficient* [11]. *Silhouette Coefficient* adalah metode yang menggabungkan teknik pemisahan dan teknik kekompakan untuk memvalidasi klaster [12].

Tahapan perhitungan Silhouette coefficient adalah sebagai berikut [13]:

1. Hitung rata-rata jarak dari suatu data misalkan  $i$  dengan semua data lain yang berada dalam satu klaster ( $a_i$ ).

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j)$$

Dengan  $j$  adalah data lain dalam satu cluster  $A$  dan  $d(I, j)$  adalah jarak antara data  $I$  dengan  $j$ .

2. Hitung rata-rata jarak dari data  $i$  tersebut dengan semua data diklaster lain, dan diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|A|-1} \sum_{j \in C} d(i, j)$$

Dengan  $d(I, C)$  adalah jarak rata-rata data  $I$  dengan semua objek pada cluster lain  $C$  dimana  $A \neq C$ .

$$b(i) = \min_{C \neq A} d(i, C)$$

3. Rumus Silhouette coefficient adalah:

$$s(i) = (b(i) - a(i)) / \max(a(i), b(i))$$

Dengan  $s(i)$  adalah semua rata-rata pada semua kumpulan data.

Untuk menganalisis hasil validasi klaster menggunakan *Average Silhouette width* dapat digunakan untuk menganalisis hasil validasi klaster, informasi tentang rentang nilai dan cara interpretasinya tersedia dalam Tabel 2.1 [14].

Tabel 1. Rentang Nilai *Silhouette*

Rentang Nilai	Interpretasi
0.71 – 1.0	Struktur yang kuat
0.51 – 0.70	Struktur yang Baik
0.25 – 0.50	Struktur yang Lemah
<0.25	Tidak ditemukan struktur yang substansial

### 2.3 HIV

*Human Immunodeficiency Virus* atau yang biasa kita sebut sebagai HIV adalah penyakit yang menyebabkan sistem kekebalan tubuh rusak. Virus ini mampu menyerang sel-sel yang ada di dalam tubuh manusia, tetapi sel yang sangat rentan untuk dirusak adalah limfosit CD4, sel tersebut juga biasa disebut Sel CD4, sel T-pembantu, atau sel pembantu. Sel CD-4 adalah jenis sel darah putih yang bertugas mengontrol dan mencegah infeksi oleh berbagai parasit, virus, jamur, dan bakteri, serta sejumlah jenis kanker. Infeksi HIV mengakibatkan kerusakan pada sel-sel CD4. Seiring berjalannya waktu, jumlah sel CD4 dapat mengalami penurunan, meskipun ini mungkin memakan waktu bertahun-tahun. Akhirnya, jumlah CD4 menjadi sangat rendah sehingga tidak lagi mencukupi untuk melawan infeksi, yang kemudian menyebabkan timbulnya gejala atau komplikasi pada individu [15].

Dalam buku saku HIV, AIDS, dan IMS Kementerian Kesehatan Republik Indonesia menulis bahwa HIV dapat ditularkan melalui

beberapa cara, karena virus ini terdapat didalam darah dan juga cairan tubuh yang lain, termasuk air susu ibu, cairan vagina, dan cairan sperma. Penyebaran HIV dapat terjadi melewati:

1. Hubungan Seksual.

Kontak dengan cairan tubuh atau darah orang yang positif HIV saat berhubungan seksual bisa menyebabkan penularan. Risiko penularan meningkat dengan frekuensi pergantian pasangan atau hubungan seksual tanpa penggunaan kondom. Jenis hubungan seksual meliputi seks vaginal (melalui vagina), seks oral (melalui mulut), dan seks anal (melalui anus). Berbagai penggunaan secara bersama suntikan, alat tato, dan alat tindik yang terkontaminasi. Jika terjadi penggunaan bersama alat ini yang terkontaminasi virus HIV, dapat mengakibatkan risiko penularan infeksi kepada individu yang terlibat dalam praktik tersebut.

2. Penularan virus dari ibu kepada bayi

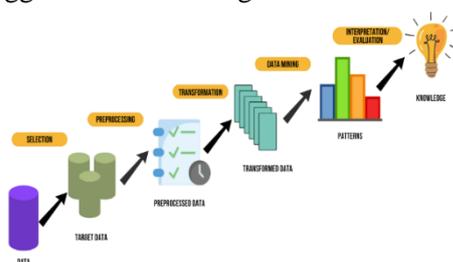
Ibu yang memiliki infeksi HIV bisa menyebabkan penularan virus ini kepada bayi selama masa kehamilan, persalinan, atau periode menyusui.

3. Pemberian darah atau produk darah lain melalui transfusi.

Proses penularan virus HIV dapat terjadi melalui pemberian transfusi darah atau produk darah lainnya yang telah terkontaminasi, sehingga orang yang menerima transfusi tersebut berisiko terinfeksi.

### 3. METODE PENELITIAN

Metodologi yang digunakan pada penelitian ini adalah KDD (Knowledge Discovery in Database). Metodologi ini melibatkan serangkaian tahap yang terstruktur dan terurut untuk mengolah data mentah menjadi informasi yang berarti dan berguna. Berikut ini adalah beberapa tahapan yang akan dilakukan dengan menggunakan metodologi KDD:



Gambar 2. Metode Penelitian

### 3.1 Data Selection

Sebelum tahap penggalian informasi dalam KDD dimulai, proses pemilihan atau seleksi data dari rangkaian data operasional dilakukan. Pada tahap ini, data yang dianggap diperlukan untuk proses akan dipilih.

### 3.2 Data Preprocessing

Data yang telah didapat, baik yang berasal dari basis data perusahaan atau hasil dari percobaan, sering kali ditemukannya ketidaksempurnaan seperti data yang salah, keberadaan data yang hilang, atau kesalahan penulisan. Selain itu, hipotesis *data mining* yang dimiliki mungkin tidak sesuai dengan beberapa atribut yang ada didalam data tersebut. Oleh karena itu, data yang dianggap tidak sesuai akan dihilangkan untuk mencegah hasil data mining yang kurang berkualitas atau akurat.

### 3.3 Data Transformation

Sebelum dapat digunakan, beberapa teknik data mining membutuhkan format data tertentu. Metode klastering dan analisis asosiasi hanya bisa memproses data numerik. Oleh karena itu, data kategorikal perlu ditransformasikan menjadi interval tertentu, yang dikenal sebagai binning. Selain itu, penyeleksian data yang relevan dengan metode data mining juga dilakukan pada tahap ini. Pada tahap ini, transformasi dan penyeleksian data memengaruhi kualitas hasil data mining, dikarenakan ciri-ciri masing-masing metode pemrosesan data bergantung pada proses ini.

### 3.4 Data Mining

Data mining adalah tahap di mana berbagai teknik, metode, atau algoritma digunakan untuk menemukan pola atau informasi yang menarik dalam data yang telah dipilih. Tujuan dan proses KDD secara keseluruhan menentukan teknik yang digunakan untuk data mining.

### 3.5 Evaluation/ Knowledge

Pada tahap Evaluasi, pola-pola yang telah diidentifikasi oleh sistem diterjemahkan menjadi pengetahuan yang bisa dimengerti oleh pemangku kepentingan.

## 4. HASIL DAN PEMBAHASAN

Hasil dari penelitian yang telah dilaksanakan merupakan hasil analisis mendalam mengenai penyebaran penyakit *Human Immunodeficiency Virus* (HIV) di Kabupaten Karawang.

Hasil penelitian ini disusun ke dalam beberapa kelompok klaster menggunakan metode *K-means Clustering*.

#### 4.1 Data Selection

Pada tahap *Data Selection*, akan dilakukan seleksi data yang di peroleh dari Dinas Kesehatan Kabupaten Karawang. *Dataset* yang diperoleh dari Dinas Kesehatan Kabupaten Karawang adalah mengenai penyakit HIV yang terdapat di Kabupaten Karawang dalam rentang waktu 2020-2023, dengan atribut nama puskesmas maupun rumah sakit, target, capaian, %, dan HIV+. Tidak semua atribut akan diolah menggunakan *K-means*; hanya atribut Puskesmas, capaian (yang akan diubah namanya menjadi Diperiksa), dan HIV+. Atribut target & % akan dihilangkan, karena atribut Puskesmas dan HIV+ lebih relevan untuk penelitian ini. Selain itu, akan ditambahkan atribut Letak untuk mengetahui lokasi kecamatan puskesmas dan rumah sakit. Untuk penelitian kali ini, akan dilakukan juga penggabungan pada kolom jumlah data HIV dari tahun 2020 – 2023. Hasil dari *Data Selection* terdapat pada Gambar 3.

NO	PUSKESMAS	LETAK	2020 - 2023	2020 - 2023
			Diperiksa	HIV +
1	Adiarsa	Karawang Timur	3754	13
2	Anggadita	Klari	1204	2
3	Balongsari	Rawamerta	1596	3
4	Batujaya	Batujaya	4272	34
5	Bayurlor	Cilamaya Kulon	1357	7
6	Ciampel	Ciampel	3686	68
7	Cibuaya	Cibuaya	1394	7
8	Cicinde	Banyusari	1654	10
9	Cikampek	Cikampek	6371	67
10	Cikampek Utara	Cikampek	1458	9
...	...	...	...	...
...	...	...	...	...
77	RS Hastien	Rengasdengklok	0	0

Gambar 3. Data Selection

#### 4.2 Data Preprocessing

Pada tahap data preprocessing akan dilakukan pembersihan data untuk mengatasi missing values, dan juga duplikasi data. Apabila terdapat data yang hilang atau sama persis maka akan diatasi agar dapat dilanjutkan ke tahapan selanjutnya.

Gambar 4 menunjukkan semua atribut, termasuk Puskesmas, Letak, Diperiksa, dan

HIV+, tidak memiliki missing values, menandakan data lengkap tanpa kekosongan.

```
[5]: #Melakukan Pengecekan Missing Values
data_select.isnull().sum()

[5]: Puskesmas    0
Letak            0
Diperiksa       0
HIV +           0
dtype: int64
```

Gambar 4. Hasil Cek Missing Values

Selain pengecekan missing values, juga akan dilakukan deteksi duplikasi data pada setiap atribut untuk memastikan setiap data unik dan tidak ada duplikasi yang bisa mempengaruhi hasil analisis.

```
[6]: #Melakukan pengecekan Duplikasi Data
data_select.duplicated().sum()

[6]: 0
```

Gambar 5. Hasil Cek Duplikasi Data

Pada Gambar 5 memperlihatkan hasil dari pengecekan duplikasi data, yang mana didalam dataset tersebut, tidak terdapat duplikasi data sama sekali.

Setelah melakukan pengecekan missing values dan juga duplikasi data, dapat disimpulkan bahwa data yang akan diolah nantinya berstatus baik dan dapat dilakukan tahapan selanjutnya.

#### 4.3 Data Transformation

Tahap Data Transformation mengubah tipe data kategorikal menjadi numerik untuk mempersiapkan data bagi algoritma *K-Means Clustering* yang memerlukan format numerik. Pada tahap ini, atribut letak yang awalnya kategorikal akan diubah menjadi numerik.

LETAK	INISIALISASI
Banyusari	1
Batujaya	2
Ciampel	3
Cibuaya	4
Cikampek	5
Cilamaya Kulon	6
Cilamaya Wetan	7
Cilebar	8
Jatisari	9
Jayakarta	10
Karawang Barat	11
Karawang Timur	12
Klari	13
Kota Baru	14
Kutawaluya	15
Lemahabang	16
Majalaya	17
Pakisjaya	18
Pangkalan	19
Pedes	20
Purwasari	21
Rawamerta	22
Rengasdengklok	23
Tegalvaru	24
Telagasari	25
Telukjambe Barat	26
Telukjambe Timur	27
Tempuran	28
Tirtajaya	29
Tirtamulya	30

Gambar 6. Hasil Inisialisasi

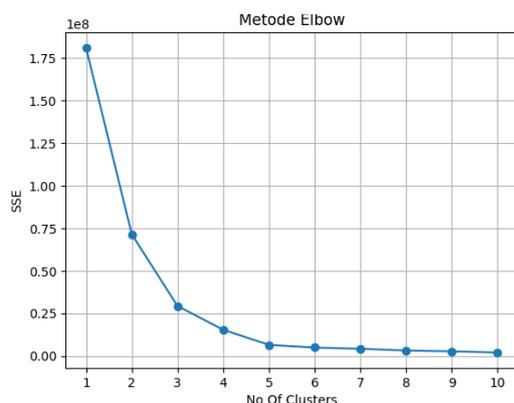
Proses inialisasi pada atribut letak melibatkan 30 angka inialisasi yang digunakan pada masing-masing nama Kecamatan. Setelah proses inialisasi selesai dilakukan, hasil dari proses *data transformation* dapat dilihat pada gambar 7.

Letak	Diperiksa	HIV +
12	3754	0
13	1204	0
22	1596	2
2	4272	8
6	1357	0
3	3686	20
4	1394	0
1	1654	0
5	6371	10
...	...	...
...	...	...
23	0	0

Gambar 7. Hasil Transformasi Data

#### 4.4 Data Mining

Tahapan yang selanjutnya adalah tahap Data Mining. Algoritma yang akan digunakan pada tahap ini adalah K-Means *Clustering* untuk mengidentifikasi daerah-daerah rawan HIV di wilayah Kabupaten Karawang. Untuk mendapatkan cluster optimal, sebelum K-Means diterapkan, akan digunakan metode elbow dengan menghitung Sum of Square Error (SSE) dari masing-masing kluster. Proses menemukan cluster terbaik dapat dilihat pada Gambar 8.



Gambar 8. Grafik Metode Elbow

Pada Gambar 8 terlihat bahwa grafik elbow untuk dataset HIV di Kabupaten Karawang pada kluster 3 jumlah nilai SSE sudah mulai stabil dan tidak mengalami penurunan yang cukup signifikan pada kluster berikutnya, dan

juga pada kluster 3 membentuk gambar siku, sehingga mendapatkan jumlah kluster yang paling optimal adalah 3.

Tahapan selanjutnya setelah mengetahui jumlah kluster optimal yaitu 3 kluster adalah proses klustering dengan menerapkan algoritma K-Means. Proses penerapan algoritma K-Means dapat dilihat pada Gambar 9 dibawah ini.

```
[12]: #Mengaplikasikan K-Means
km = KMeans(n_clusters=3, init='k-means+', random_state=42)
km.fit(data_transform)
centroid = km.cluster_centers_
centroid

[12]: array([[ 14.83333333, 1899.76190476, 26.95238095],
 [ 15.54166667, 211.25, 5.25],
 [ 14.09090909, 4672.72727273, 29.18181818]])
```

Gambar 9. Hasil Perhitungan dengan K-Means

Pada gambar 9 terlihat bahwa kluster 0 menunjukkan nilai *centroid* berada ditengah-tengah antara kluster 1 & 2 ini menunjukkan bahwa kluster 0 adalah kluster dengan kriteria rawan. Kluster 1 berada di kriteria tidak rawan, dan kluster 2 sangat rawan.

Untuk mengetahui jumlah anggota dari masing-masing kluster, maka dilakukan proses mendreskipsikan masing-masing kluster, seperti yang ada pada Gambar 10.

```
[13]: #melihat isi dari masing2 cluster
data_transform['Cluster'] = km.labels_
n_cs = 3
for i in range(n_cs):
    cnt = data_transform[data_transform['Cluster'] == i]['Cluster'].count()
    print(f'Cluster {i+0}: {cnt}')

Cluster 0: 42
Cluster 1: 24
Cluster 2: 11
```

Gambar 10. Anggota Masing-masing Kluster

Pada Gambar 10 terlihat bahwa data set HIV di Kabupaten Karawang dengan pengklasteran menggunakan jumlah 3 kluster, untuk kluster 0 terdapat 42 anggota, kluster 1 terdapat 24 anggota, dan kluster 2 terdapat 11 anggota. Kluster 0 mendapatkan anggota terbanyak dibandingkan kluster yang lainnya.

#### 4.5 Evaluation/ Knowledge

Tahap terakhir adalah Evaluasi/ Interpretasi. Tahapan ini sangat penting dilakukan, karena hasil dari analisis yang telah diperoleh harus dievaluasi untuk menilai sejauh mana kualitas analisis yang diterapkan sesuai dengan tujuan penelitian. Pada tahapan ini akan digunakan evaluasi dengan metode Silhouette Coefficient.

```
[18]: #Melakukan evaluasi dengan silhoutte coefficient
from sklearn.metrics import silhouette_score

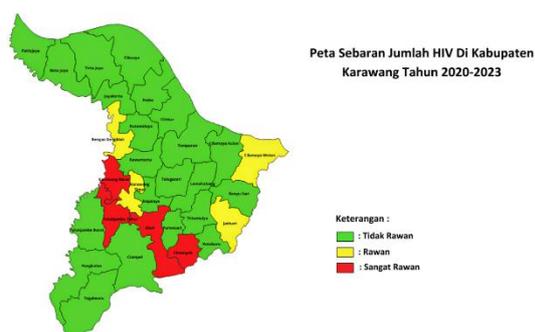
shill_avg = silhouette_score(data_transform, km.labels_)
print('Nilai Silhoutte Coefficient: '+str(shill_avg))

Nilai Silhoutte Coefficient: 0.6343444342900506
```

**Gambar 11.** Hasil Proses Evaluasi

Pada Gambar 11 terlihat hasil dari nilai Silhouette Coefficient adalah 0.6343 yang mana hasil tersebut menunjukkan bahwa kluster yang didapatkan baik atau bisa disebut dengan *Medium Structure* atau tersebut dapat dikatakan memiliki kualitas kluster yang baik.

Setelah selesai melakukan evaluasi, selanjutnya akan dilakukan visualisasi pemetaan menggunakan tools QGIS. Dalam proses visualisasi menggunakan tools QGIS, kluster tersebut direpresentasikan dengan 3 warna berbeda, yakni hijau, kuning, dan merah. Warna hijau digunakan untuk menandai daerah yang tidak rawan terhadap penderita HIV, sedangkan warna kuning menunjukkan daerah yang memiliki tingkat rawan sedang, dan warna merah mengindikasikan daerah yang memiliki tingkat rawan tinggi atau jumlah penderita HIV yang signifikan. Dengan demikian, visualisasi ini memungkinkan untuk dengan cepat dan jelas memahami distribusi spasial dari penyakit HIV di Kabupaten Karawang serta tingkat risiko yang terkait di setiap daerah. Untuk visualisasi peta sebaran HIV di setiap kecamatan terdapat pada Gambar 12.



**Gambar 12.** Hasil Visualisasi dengan Qgis

## 5. KESIMPULAN

- Penerapan algoritma K-Means untuk mengelompokkan penyakit HIV di Kabupaten Karawang tahun 2020-2023 berhasil dengan baik. Penelitian ini menghasilkan 3 kluster berdasarkan

tingkat kerentanan: Kluster 1 Tidak Rawan (24 anggota), Kluster 0 Rawan (42 anggota), dan Kluster 2 Sangat Rawan (11 anggota).

- Setelah menerapkan metode elbow, diperoleh hasil terbaik dengan 3 kluster. Grafik metode elbow menunjukkan titik siku yang memperlihatkan bahwa pembagian dataset menjadi 3 kluster adalah yang paling optimal.
- Hasil evaluasi penerapan algoritma K-Means untuk pengelompokkan penyakit HIV di Kabupaten Karawang tahun 2020-2023 menggunakan Silhouette Coefficient mendapatkan nilai 0,63, menunjukkan bahwa kluster yang terbentuk memiliki struktur yang baik.
- Hasil visualisasi pemetaan penyebaran HIV di Kabupaten Karawang tahun 2020-2023 menggunakan QGIS menunjukkan 3 kluster dengan warna berbeda sesuai tingkat kerentanan. Kluster sangat rawan (merah) meliputi Kecamatan Cikampek, Telukjambe Timur, Karawang Barat, dan Klari. Kluster rawan (kuning) mencakup Kecamatan Jatisari, Karawang Timur, Rengasdengklok, dan Cilamaya Wetan. Kluster tidak rawan (hijau) mencakup Kecamatan Banyu Sari, Batu Jaya, Ciampel, Cibuaya, Cilamaya Kulon, Cilebar, Jayakarta, Kotabaru, Kutawaluya, Lemahabang, Majalaya, Pakisjaya, Pangkalan, Pedes, Purwasari, Rawamerta, Tegalwaru, Talagasari, Telukjambe Barat, Tempuran, Tirta Jaya, dan Tirtamulya.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini dan telah membantu penulis untuk menyelesaikan penelitian ini.

## DAFTAR PUSTAKA

- H. Noor, A. Dharmawati, and T. W. Qur'ana, "Penerapan Algoritma K-Means Clustering Analysis Pada Kasus Penderita Hiv/Aids (Studi Kasus Kabupaten Banjar)," *Technol. J. Ilm.*, vol. 12, no. 2, p. 72, 2021, doi: 10.31602/tji.v12i2.4573.

- [2] D. Rohmatullailah and D. Fikriyah, "Faktor Risiko Kejadian HIV Pada Kelompok Usia Produktif di Indonesia," *J. Biostat. Kependudukan, dan Inform. Kesehat.*, vol. 2, no. 1, p. 45, 2021, doi: 10.51181/bikfokes.v2i1.4652.
- [3] R. I. L. Sinaga, W. Saputra, and H. Qurniawan, "Pengelompokan Jumlah Kasus Penyakit Aids Berdasarkan Provinsi Menggunakan Metode K-Means," *J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 2, pp. 99–107, 2021.
- [4] D. Y. Nugroho, A. Wibowo, F. Kesehatan, and M. Universitas, "ANALISIS CLUSTER K - MEANS KABUPATEN / KOTA DI JAWA TIMUR BERDASARKAN LEVEL EPIDEMIK HIV CLUSTER K - MEANS ANALYSIS OF THE HIV EPIDEMIC LEVEL IN," vol. 8, no. 2, pp. 108–117, 2019, doi: 10.20473/jbk.v8i2.2019.108.
- [5] Z. Khan, "HIV-INFECTED PEOPLE DATA," vol. 100, no. 13, pp. 5010–5018, 2022.
- [6] N. A. Mufid, D. Risqi, N. Hanum, and A. H. Sidiq, "Clustering Kabupaten / Kota di Jawa Tengah Tahun 2022 berdasarkan Jumlah Kasus Kemunculan Penyakit dengan Algoritma K-Means," no. 1, pp. 1–14, 2023.
- [7] F. Novianti, Y. R. Aisyah Yasmin, and D. C. R. Novitasari, "Penerapan Algoritma Fuzzy C-Means (FCM) dalam Pengelompokan Provinsi di Indonesia berdasarkan Indikator Penyakit Menular Manusia," *JUMANJI (Jurnal Masy. Inform. Unjani)*, vol. 6, no. 1, p. 23, 2022, doi: 10.26874/jumanji.v6i1.103.
- [8] T. P. Sari, A. L. Hananto, E. Novalia, T. Tukino, and S. S. Hilabi, "Implementasi Algoritma K-Means dalam Analisis Klasterisasi Penyebaran Penyakit Hiv/Aids," *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 104–114, 2023, doi: 10.29408/jit.v6i1.7423.
- [9] E. M. Fitri, R. R. Suryono, and A. Wantoro, "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz," *J. Komputasi*, vol. 11, no. 2, pp. 157–168, 2023, doi: 10.23960/komputasi.v11i2.12582.
- [10] W. Rianti and E. Harahap, "Pengolahan Data Hasil Penjualan Online Menggunakan Aplikasi Microsoft Excel," *Mat. J. Teor. dan Terap. Mat.*, vol. 20, no. 2, pp. 69–76, 2021, [Online]. Available: <https://journals.unisba.ac.id/index.php/matematika/article/view/1553>
- [11] P. S. Rosiana, A. A. Mohsa, and Y. Umidah, "Implementasi K-Means Dalam Pengelompokan Penyebaran Penyakit Dbd Di Jawa Barat," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, pp. 782–788, 2023, doi: 10.23960/jitet.v11i3.3344.
- [12] D. I. Yunistya, R. Goejantoro, F. Deny, and T. Amijaya, "The Application Of K-Harmonic Means Method In District/City Grouping (Case Study: Poverty in Kalimantan Island in 2020) Penerapan Metode K-Harmonic Means Dalam Pengelompokan Kabupaten/Kota (Studi Kasus: Kemiskinan di Pulau Kalimantan Tahun 2020)," *J. Mat. Stat. dan Komputasi*, vol. 19, no. 1, pp. 51–64, 2022, doi: 10.20956/j.v19i1.21116.
- [13] A.- Shoolihah, M. T. Furqon, and A. W. Widodo, "Implementasi Metode Improved K-Means untuk Mengelompokkan Titik Panas Bumi," vol. 1, no. 11, pp. 1270–1276, 2019.
- [14] B. N. Sari and A. Primajaya, "Penerapan Clustering DbSCAN Untuk Pertanian Padi Di Kabupaten Karawang," *J. Inform. dan Komput.*, vol. 4, no. 1, pp. 28–34, 2019, [Online]. Available: <https://ejournal.akakom.ac.id/index.php/jiko/article/view/178%0Awww.mapcoordinates.net/en>
- [15] N. Kristiono and I. Astuti, "Mengenal HIV & AIDS," *Https://Www.Researchgate.Net/Publication/337049851*, no. June, p. 80, 2019, [Online]. Available: [https://r.search.yahoo.com/\\_ylt=AwrX.HbEF99k5iQFLgRP5At.;\\_ylu=Y29sbwMEcG9zAzUEDnRpZAMEc2VjA3Ny/RV=2/RE=1692371012/RO=10/RU=https%3A%2F%2Fwww.who.int%2Fnews-room%2Ffact-sheets%2Fdetail%2Fhiv-aids/RK=2/RS=bB8wb4.Rwts28NCVTQ1vv31uJ4Y-](https://r.search.yahoo.com/_ylt=AwrX.HbEF99k5iQFLgRP5At.;_ylu=Y29sbwMEcG9zAzUEDnRpZAMEc2VjA3Ny/RV=2/RE=1692371012/RO=10/RU=https%3A%2F%2Fwww.who.int%2Fnews-room%2Ffact-sheets%2Fdetail%2Fhiv-aids/RK=2/RS=bB8wb4.Rwts28NCVTQ1vv31uJ4Y-)