

ANALISIS CLUSTER PADA KELOMPOK MASYARAKAT YANG RENTAN TERHADAP PAPARAN COVID-19 MENGGUNAKAN METODE K-MEANS CLUSTERING DAN VISUALISASI DENGAN SIG

Indra Raja Drl ¹, Yulison Herry Chrisnanto², Fajri Rakhmat Umbara ³

^{1,2,3} Jurusan Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia
e-mail: indrarajadrl17@if.unjani.ac.id¹, y.chrisnanto@gmail.com², fajri.umbara@gmail.com³

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi : 01 Februari 2022

Revisi Akhir : 13 Juni 2022

Diterbitkan Online : 30 November 2022

Kata Kunci :

K-Means, Clustering, Covid-19, Sig, Visualisasi

Korespondensi :

Telepon / Hp : +62 89614730295

E-mail : indrarajadrl17@if.unjani.ac.id

ABSTRAK

Covid-19 adalah penyakit yang menular serta laju infeksi yang cepat, setelah mencapai 100 kasus yang dikonfirmasi terinfeksi tingkat penyebarannya meluas. Dengan cepatnya penyebaran wabah Covid-19 masyarakat sangat prihatin dengan penyebaran dan dampaknya, orang yang sebelumnya sudah memiliki gangguan kesehatan akan meningkatkan risiko terinfeksi Covid-19 gangguan kesehatan ini seperti, tuberkulosis, diabetes, diare, hipertensi. Adapun Faktor lain yang mempengaruhi penyebaran Covid-19 seperti kepadatan penduduk yang tinggi di kota besar, iklim, suhu dan daerah metropolitan merupakan faktor risiko utama untuk tertular virus. Dari adanya faktor yang mempengaruhi kasus Covid-19 sehingga Satgas Penanganan Covid-19 menilai pentingnya bagi semua pihak termasuk masyarakat memahami faktor-faktor lonjakan kasus Covid-19 agar terhindar dari kasus itu. Tujuan dari penelitian ini menggunakan metode K-Means Clustering untuk analisis cluster pada wilayah yang memiliki karakteristik tingginya kasus Covid-19 dan variabel apa yang berpengaruh terhadap tingginya kasus Covid-19 dan visualisasi menggunakan Sistem Informasi Geografis sehingga diharapkan dapat menjadi informasi bagi masyarakat dan instansi kesehatan untuk memahami kelompok wilayah yang rentan. Kesimpulannya wilayah kota Bandung dikelompokkan menjadi 3 cluster yang dimana cluster 1 itu wilayah dengan kasus Covid-19 tertinggi dan faktor yang mempengaruhi Covid-19 juga tinggi untuk cluster 2 memiliki tingkat kasus yang rendah dan cluster 3 memiliki tingkatan yang lebih rendah dari kedua cluster.

1. PENDAHULUAN

Penyakit baru Covid-19 telah menimbulkan bahaya yang luar biasa bagi masyarakat. Covid-19 adalah penyakit yang menular serta laju infeksi yang cepat. Pada tahap awal jumlah terinfeksi kasus sedikit meningkat, namun, setelah mencapai 100 kasus yang dikonfirmasi terinfeksi, tingkat pertumbuhannya cepat atau sedikit ada perubahan sehingga WHO menyatakan wabah penyakit di Cina Covid-19 menjadi Darurat Kesehatan Masyarakat karena menimbulkan risiko tinggi untuk terpapar bagi negara-negara dengan sistem kesehatan yang kurang [1], [2], [3]. seperti di negara Amerika Serikat menjadi negara terdepan jumlah kasus covid di seluruh dunia, menggantikan Italia yang sebelumnya masuk memimpin kasus Covid-19 [4]. dampak adanya kasus COVID-19 ini di Thailand lebih dari 10 jutaan orang akan menghadapi pengangguran diakibatkan oleh adanya pembatasan [2].

Dengan cepatnya penyebaran wabah Covid-19, masyarakat sangat prihatin dengan penyebaran dan

dampaknya [5]. orang yang sebelumnya sudah memiliki gangguan kesehatan akan meningkatkan resiko terinfeksi Covid-19 [6]. gangguan kesehatan ini dikaitkan dengan Penyakit tuberkulosis dengan peningkatan 2,10 kali lipat risiko penyakit Covid-19 yang parah. tuberkulosis memiliki peluang lebih tinggi untuk mengembangkan komplikasi serius dari Covid-19 [7]. Diabetes melitus dikaitkan perkembangan penyakit pada pasien dengan Covid-19 [8] adapun masyarakat yang didiagnosis memiliki obesitas menjadi pengaruh terinfeksi Covid-19 dan membawanya lebih parah [9]. Saat pasien memiliki diare, merupakan gejala yang dapat terkonfirmasi karena pandemi Covid-19 [10]. Adapun Faktor lain yang mempengaruhi penyebaran Covid-19 seperti kepadatan penduduk yang tinggi di kota besar, iklim, suhu dan daerah metropolitan merupakan faktor risiko utama untuk tertular virus. [11].

Dari adanya faktor yang mempengaruhi kasus Covid-19 sehingga Satgas Penanganan code menilai pentingnya bagi semua pihak termasuk masyarakat memahami

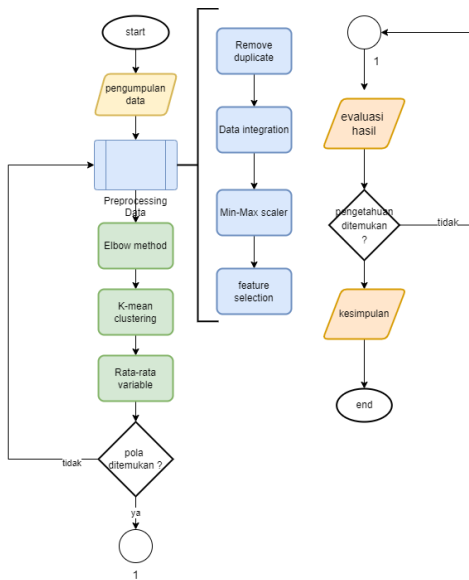
faktor-faktor lonjakan kasus Covid-19 agar terhindar dari kasus itu[12].

Clustering metode yang biasa digunakan untuk memetakan atau mengklasifikasikan suatu objek adalah algoritma K-Means[13]. pada penelitian lainnya menggunakan metode k-mean clustering dapat digunakan untuk menentukan indikator yang mempengaruhi penyebaran Covid-19[14]. Sebuah penelitian yang dilakukan di amerika serikat yaitu GIS-COVID19 mengelompokkan 3 cluster, Mereka adalah cluster wilayah dengan tingkat pendapat yang tidak sama, umr yang rendah dan partisipasi perawat. Faktor risiko ini berpengaruh signifikan dalam hasil COVID-19[15].

Maksud dari penelitian ini adalah analisis cluster Menggunakan metode K-Means Clustering sehingga didapat hasil cluster dan menentukan karakteristik cluster dengan dilakukannya perbandingan untuk setiap cluster, setelah itu divisualisasi sehingga menjadi informasi bagi masyarakat dan instansi kesehatan untuk memahami kelompok wilayah mana yang memiliki jumlah kasus Covid-19 tertinggi dan faktor apa yang mempengaruhi tingginya kasus Covid-19 pada kelompok itu.

2. METODE PENELITIAN

Terdapat tahapan-tahapan yang dilakukan dalam penelitian ini yang digambarkan dengan flowchart pada gambar 1.



Gambar 1. Metode Penelitian

A. Pengumpulan Data

dalam tahapan ini mengambil data komorbid dan populasi yang didapat dari situs open data dinas kesehatan kota bandung, data tersebut mengenai jenis-jenis penyakit dan jumlah penduduk pada seluruh kecamatan.

B. Preprocessing Data

Preprocessing data mencakup antara lain membuang duplikasi data, melakukan integrasi data, merubah skala data dan seleksi fitur pada data, tujuannya agar data yang digunakan mendapatkan hasil informasi yang sesuai dan juga format yang baik tergantung terhadap metode yang digunakan.

1. *Remove duplicate*, Menghapus duplikasi data yang bertujuan agar data yang digunakan tidak memiliki kesamaan sehingga lebih sesuai.
2. *Data integration*, Menggabungkan data yang sama sehingga dalam satu format, yang sebelumnya data itu terpisah karena dari source yang berbeda.
3. *Min-max scaler*, Mengubah skala data yang tinggi, manfaatnya data memiliki skala yang lebih rendah dengan rentang yang ditentukan semisalnya angka minimal 0 dan maksimalnya 2.

Rumus : menghitung min-max scaler

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Keterangan :

- Min : minimum
- Max : maximum
- X : value range

4. *Selection feature*, Seleksi variabel yang digunakan, jika variabel yang didapat tidak mengandung sebuah informasi maka variabel tersebut akan dihapus.

C. Elbow Method

Metode ini berfokus pada persentase varian sebagai fungsi jumlah cluster. Nilai dari k ditambahkan satu per satu dan nilai Sum Square Error (SSE) direkam[16]. metode elbow ini digunakan untuk menentukan jumlah k optimal dari jumlah data yang ada.

Rumus : Menghitung SSE

$$SSE = \sum_{k=1}^k \sum_{xi \in sk} ||Xi - Ck||^2 \tag{2}$$

Keterangan :

- SSE : Sum of Square Error
- Xi : nilai atribut dari data ke-i
- Ck : nilai atribut titik pusat cluster ke-I
- K : Banyaknya cluster

Saat nilainya turun secara drastis dan membentuk sudut yang lebih kecil (membentuk seperti siku-siku tangan), maka nilai k optimalnya telah ditemukan. semakin banyak jumlah datanya maka tingkat SSEnya semakin tinggi.

D. K-Means Clustering

K-Means adalah metode pengelompokan yang cepat dan sederhana dengan jumlah iterasi yang lebih

kecil. Algoritma ini membagi data menjadi bagian k. Persyaratan *cluster* diperkirakan berdasarkan pengguna pilihan. Komputer secara acak memilih dan menetapkan objek ke satu objek *cluster* (k). Jarak antara setiap objek dan pusat setiap *cluster* dihitung dan menghasilkan *cluster* yang optimal larutan[16].

Rumus : Hitung jumlah jarak kuadrat antara titik data dan semua centroid.

$$d_{i,j} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \quad (3)$$

Keterangan :

- Dij : Jarak objek antara objek i dan j
- P : Dimensi data
- Xik : Koordinat dari obyek i pada dimensi k
- Xjk : Koordinat dari obyek j pada dimensi

Rumus : Hitung centroid untuk *cluster* dengan mengambil rata-rata dari semua titik data yang dimiliki setiap *cluster*.

$$\mu_k = \frac{1}{N_k} \sum_{j=1}^{N_k} x_j \quad (4)$$

Keterangan :

- μ_k : adalah titik tengah kelompok ke-I/k
- Nk : banyak data dalam *cluster* ke-I/k
- I,K : indeks dari *cluster*
- j : indeks dari variable(v1,v2,v3..)
- Xj : nilai data ke-j dalam ke-I/k

E. Rata-Rata Variable

Nilai rata-rata adalah angka representasi atau biasa juga disebut sebagai mean dari suatu kelompok data yang mewakili data secara keseluruhan. Dapat digunakan untuk membandingkan kualitas suatu kelompok dengan kelompok lain ,Mewakili kondisi suatu kelompok dengan 1 angka dan Mendeskripsikan suatu kelompok dengan singkat[17].

Rumus : menghitung rata-rata variabel

$$x = \frac{1}{n} \sum_{i=1}^n xi \quad (5)$$

Keterangan :

- X : rata-rata hitung
- Xi : nilai sample ke-I
- n : jumlah sample

F. Evaluasi Hasil

Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.Setelah teruji data bisa dipresentasikan kepada pengguna.

1. *silhouette coefficient* berfungsi untuk menguji kualitas dari *cluster* yang dihasilkan serta sebagai metode pengesahan suatu *cluster* yang menggabungkan antara metode kohesion dan

metode separation [18]. Hasil perhitungan nilai silhouette coefficient memiliki range antara -1 hingga 1. Hasil dapat dikatakan baik apabila bernilai positif, hal ini berarti titik sudah berada di dalam *cluster* yang tepat sedangkan jika nilainya negatif ini menandakan terjadinya overlapping sehingga titik berada di antara dua *cluster* .

Rumus : Menghitung silhouette coefficient

$$S_i = \frac{b_i - a_i}{\max(a_i - b_i)} \quad (5)$$

Keterangan :

- Si : nilai silhouette coefficient
- ai : rata-rata jarak titik *cluster* n dengan semua data yang bek rada dalam *cluster* n
- bi : minimal jarak rata-rata dari titik *cluster* n ke titik *cluster* yang lain yang berbeda dari *cluster* n

2. *Davies-bouldin index (DBI)* adalah rasio dari jumlah jarak intra-*cluster* dan jarak inter-*cluster* . Untuk mencapai hasil *clustering* yang baik, jarak inter-*cluster* harus tinggi dan jarak intra-*cluster* harus rendah dan oleh karena itu, nilai DBI yang lebih rendah *cluster* terbaik[19].Rumus : menghitung davies-bouldin index

Rumus : menghitung DBI

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (6)$$

Keterangan :

- K : jumlah *cluster*
- R_{ij} : rasio perbandingan antara *cluster* ke-i dan *cluster* ke-j

Semakin kecil nilai Davies Bouldin Index (DBI) yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering*.

3. *Visualisasi dengan sistem informasi geografis* ,Sistem Informasi Geografis(GIS) merupakan sebuah tool yang digunakan untuk memetakan wilayah secara geografis , bisa juga digunakan untuk menampilkan tingginya prevalensi ,transmisi suatu penyakit dan memodelkan lingkungan secara spatial[20].

3. PERANCANGAN SISTEM

Pada perancangan sistem ini adalah isi dari tahapan-tahapan pada metode penelitian yang digunakan.

A. Pengumpulan Data

Data yang diperoleh dan dikumpulkan dari situs open data dinas kesehatan kota bandung , prosesnya mengambil semua data jenis-jenis penyakit dan data

kependudukan untuk dilakukan analisa pengaruhnya terhadap kasus Covid-19. data yang diperoleh sebanyak 30 dan 7 variable menyatakan kecamatan, jenis-jenis komorbid dan populasi. Data ini merupakan rekapitulasi data profil kesehatan tahun 2020 yang dimana ditahun ini terdapat pengaruh terhadap kasus Covid-19.

B. Preprocessing Data

Dalam preprocessing data adalah tahapan untuk mengolah data sehingga data itu siapa itu digunakan.

1. Remove duplicate

pada tahapan pertama apabila terdapat variabel yang duplikasi akan dihilangkan atau digabungkan jumlahnya untuk menghindari kemiripan variable.

Tabel 1. Remove Duplicate

Input	Output
Obesitas L, Obesitas P Tuberculosis L, Tuberculosis P	Obesitas dan tuberculosis

2. Data Integration

Pada tahapan kedua ini akan menggabungkan data dari source yang terpisah agar sesuai dengan format yang dapat diperhitungkan, semisalnya akan menggabungkan 2 file yaitu table hipertensi.xls dan table diabetes.xls yang didapat dari source yang berbeda.

Tabel 2. Data Integration

Input	Output
Hipertensi.xls Diabetes.xls	Data table dari kedua file

3. Min-Max Scaler

Pada tahapan ketiga ini melakukan Normalisasi min-max pada data asli. Nilai dinormalisasi dalam rentang yang diberikan dari 0 sampai 2. Manfaat normalisasi Min-Max adalah bahwa semua nilai adalah dalam kisaran tertentu.

Tahap 1: menentukan nilai maksimum dan minimum pada suatu variabel

Tabel 3. Penentuan min-max

variable	Minimum	maksimum
Obesitas	7	92

Tahap 2: Melakukan normalisasi data yang ke-i dengan rentang 0-2.

Tabel 4. Perhitungan min-max

Data ke-i tuberculosis	$z_i = (x_i - \min(x)) / (\max(x) - \min(x)) * 2$
Data ke-1	$(88 - 7) / (92 - 7) * 2 = 1.905882352941176$
Data ke -2	$(16 - 7) / (92 - 7) * 2 = 0.211764705882352$
Data ke-3	$(25 - 7) / (92 - 7) * 2 = 0.423529411764705$
...	..
Data ke-30	$(39 - 7) / (92 - 7) * 2 = 0.752941176470588$

Tahap 3 : Lakukan perhitungan untuk semua variabel dan data yang ke-i sehingga didapat hasilnya.

Tabel 5. Hasil min-max

n	Obesitas	Diabetes	Tuberculosis	Diare	Hipertensi	Jumlah penduduk	Positif covid
1	0.128	0.859	1.905	0.537	0.917	1.298	1.637
2	0.064	0.888	0.211	1.213	1.120	0.938	1.663
3	0.041	1.519	0.423	0.849	1.293	0.905	1.760
...
30	0.223	0.237	0.752	0.697	0.215	1.078	0.854

4. Feature Selection

Pada tahapan terakhir ini akan seleksi atau mengeliminasi variabel yang tidak memiliki relasi dengan Covid-19 berdasarkan hasil dari penelitian publik yang berkaitan dengan data komorbid Covid-19. variabel yang dieliminasi penyakit jiwa dan demam berdarah.

C. Elbow method

Pada tahapan ini untuk menentukan jumlah cluster yang optimal dengan Sum of square error Saat nilainya turun secara drastis dan membentuk sudut yang lebih kecil (membentuk seperti siku-siku tangan), maka nilai k optimalnya telah ditemukan. semakin banyak jumlah datanya maka tingkat SSEnya semakin tinggi.

Tahap 1: tentukan jumlah cluster, jumlah cluster ditentukan sebanyak 5, yang dimana perulangan dimulai dari jumlah K - 1. tentukan titik pusat cluster secara random untuk iterasi pertama.

Tabel 6 Centroid random 2

Centroid	Obesitas	diabetes	tuberculosis	diare	hipertensi	Jumlah Penduduk	Covid-19
Centroid 1	0.2575	0.7871	0.9926	0.8419	1.1329	1.3911	1.2394
Centroid 2	0.1252	0.2915	0.3613	0.3905	0.4684	0.5790	0.5829

Tahap 2 : melakukan perhitungan SSE untuk setiap data ke-i dengan cluster-nya

Tabel 7 Hitung SSE cluster 2

Data ke-i	Perhitungan SSE -i (xi-ck)
$\sum_{i=1}^2 \sum_{x_i \in C_k} \ x_i - c_k\ ^2$	$(0.128837209 - 0.2575872091875)^2 + (0.859891259 - 0.7871183605)^2 + (1.905882353 - 0.9926470588125)^2 + (0.53748232 - 0.8419377651875)^2 + (0.917912476 - 1.13293354175)^2 + (1.298979914 - 1.3911141199375)^2 + (1.637540453 - 1.2394822006875)^2$

$$\sum_{z=1}^2 \sum_{x_2 \in E_2} \|x_i - ck\|_2^2$$

(0.064883721 - 0.2575872091875)²
 (0.888331242 - 0.7871183605)²
 (0.211764706 - 0.9926470588125)²
 (1.21301273 - 0.8419377651875)²
 (1.120956782 - 1.13293354175)²
 (0.938602026 - 1.3911141199375)²
 (1.663430421 - 1.2394822006875)²

.. ..

$$\sum_{z=1}^2 \sum_{x_{30} \in E_2} \|x_i - ck\|_2^2$$

(0.223255814 - 0.12524916942857)²
 (0.237557507 - 0.29156957635714)²
 (0.752941176 - 0.36134453771429)²
 (0.697029703 - 0.39050313185714)²
 (0.215004077 - 0.46843086228571)²
 (1.078592211 - 0.57903540014286)²

Sehingga untuk setiap perhitungan dimulai dari data ke-i sampai jumlah data maksimal didapat hasil SSE untuk cluster 2 .

Tabel 8 Hasil SSE cluster 2

Data ke-i	Nilai Penjumlahan
1	1.1617374208125
2	1.1794952916257
..	..
30	0.64725958883863
Hasil penjumlahan total SSE dari data ke-1 sampai ke-30 didapat untuk cluster 2 = 29.357775054012	

Tahap 3 : Menaikan jumlah cluster , perhitungan SSE ditambah 1 cluster sampai dengan jumlah cluster yang telah ditentukan. Lalu menentukan centroid yang baru.

Tabel 9. Centroid Random 3

Centroid	Obesitas	diabetes	tuberculosis	diare	hipertensi	Jumlah Penduduk	Covid-19
Centroid 1	0.180	0.8286	1.2149	0.8671	1.2476	1.5324	1.2868
Centroid 2	0.3418	0.5422	0.5058	0.8347	0.6534	1.0125	0.5829
Centroid 3	0.1053	0.2929	0.3208	0.2473	0.5212	0.4915	1.0008

Tahap 4 : Melakukan Perhitungan SSE kembali, dengan jumlah centroid baru 3.

Tahap 5 : lakukan perhitungan sampai dengan jumlah cluster ke-5, sehingga didapat hasil SSE tiap cluster dan selisihnya. dapat ditetapkan jumlah cluster yang optimal adalah sebanyak 3 dengan 30 data dan 7 variable dari data , karena perpindahan dari cluster 2 ke cluster 3 memiliki penurunan sel yang signifikan.

Tabel 10. Centroid Random 3

Cluster	SSE	Selisih (SSE i - SSE -i+1)
2	29.35	29.35
3	25.67	3.680
4	21.36	4.308
5	22.07	0.704

D. K-Means Clustering

Dalam tahapan ini akan mengelompokan data yang memiliki karakteristik sama akan ditentukan pada kelompok yang mana , cara kerja k-mean untuk pengelompokan data yaitu menggunakan rumus euclidean distance melihat kedekatan data dengan centroid.

Tahap 1: Menentukan jumlah K/cluster , pada tahapan ini jumlah cluster ditentukan dari hasil metode elbow yang didapat jumlah k optimal sebanyak 3 cluster.

Tahap 2: Menentukan Centroid secara acak sebanyak dengan jumlah cluster-nya. Sehingga dapat dilakukan perhitungan kedekatan data point dengan centroidnya.

Tabel 11. Centroid Random 3

Centroid	Obesitas	diabetes	tuberculosis	diare	hipertensi	Jumlah Penduduk	positif Covid
1	0.358	0.273	2	0.772	1.047	2	0.886
2	0	0.756	0.894	2	1.752	1.539	0.906
3	0.058	0.792	0.494	0.017	1.150	1.150	0.899

Tahap 3 : Setelah diketahui centroid untuk masing-masing variabel, lalu masuk ke proses perhitungan euclidean distance untuk menentukan suatu data jarak paling dekatnya dengan centroid yang mana.

Tabel 12. Hitung jarak iterasi-1

Data ke 1 andir	Hitung Euclidean distance
$d_{1,1}$ Data 1 , centroid 1	$\sqrt{\sum_{k=1}^6 \{x_{ik} - x_{jk}\}^2}$ $\sqrt{(0.128837209 - 0.358837209)^2 + (0.859891259 - 0.273525721)^2 + (1.905882353 - 2)^2 + (0.53748232 - 0.772842999)^2 + (0.917912476 - 1.047567274)^2 + (1.298979914 - 2)^2 + (1.637540453 - 0.886731392)^2}$
$d_{1,2}$ Data 1 , centroid 2	$\sqrt{\sum_{k=1}^6 \{x_{ik} - x_{jk}\}^2}$ $\sqrt{(0.128837209 - 0)^2 + (0.859891259 - 0.756168967)^2 + (1.905882353 - 0.894117647)^2 + (0.53748232 - 0.017538897)^2 + (0.917912476 - 1.75292199)^2 + (1.298979914 - 1.539664178)^2 + (1.637540453 - 0.906148867)^2}$
$d_{1,3}$	$\sqrt{\sum_{k=1}^6 \{x_{ik} - x_{jk}\}^2}$ $\sqrt{(0.128837209 - 0.058372093)^2 + (0.859891259 - 0.792137181)^2 + (1.905882353 - 0.494117647)^2}$

$$\begin{aligned} \text{Data 1, centroid 3} &+ (0.53748232 - 0.928429986)^2 \\ &+ (0.917912476 - 0.017538897)^2 \\ &+ (1.298979914 - 1.150856211)^2 \\ &+ (1.637540453 - 0.899676375)^2 \end{aligned}$$

Sehingga didapat centroid untuk data 1 ,hasilnya pada tabel berikut, dari jarak terdekat untuk data 1 adalah centroid 1.

Tabel 13. Hasil jarak iterasi-1

Data ke 1 andir	Nilai centroid
Data 1 centroid 1	1.23811594713
Data 1 centroid 2	2.1166215014451
Data 1 centroid 3	1.7638958576596

Tahap 4 : melakukan perhitungan untuk setiap data dengan centroid terdekat dan menentukan termasuk kepada *cluster* yang mana ,sehingga didapat hasilnya untuk iterasi pertama.

Tabel 14. Hasil cluster iterasi-1

No	Kecamatan	Cluster
1	Andir	1
2	Antapani	3
3	Arcamanik	3
4	Astanaanyar	3
.	.	.
30	Ujung Berung	3

Tahap 5 : Penentuan Centroid baru pertama,penentuan centroid ini dilakukan agar menyesuaikan data dengan centroid yang sebenarnya.Dilakukan penentuan centroid sampai dengan tidak ada perubahan centroid .

Tabel 15. Penentuan centroid baru iterasi-1

Centroid	Penentan centroid
Centroid ke 1	$\left(\frac{1.637540453+2+0.886731392+0.705501618+0.912621359}{5} \right)$ $\approx 0.26424418575, (\text{obesitas})$ $\left(\frac{0.859891259+0.273525721+0.60811376+0.403178586+0.649100795}{5} \right)$ $\approx 0.54642409025, (\text{diabetes})$ $\left(\frac{1.905882353+1.341176471+2+0.403178586+0.649100795}{5} \right)$ $\approx 1.6470588235 (\text{tuberculosis})$ $\left(\frac{0.53748232+0.772842999+0.540311174+0.704950495+1.14115983}{5} \right)$ $\approx 0.789108911 (\text{diare})$ $\left(\frac{0.917912476+1.047567274+0.472139168+0.877684153+1.012503398}{5} \right)$ $\approx 0.96391682525 (\text{hipertensi})$ $\left(\frac{1.298979914+2+0.611052687+1.722981737+1.064570407+1.893101272}{5} \right)$ $\approx 1.5216330145 (\text{jumlah penduduk})$ $\left(\frac{1.637540453+2+0.886731392+0.705501618+0.912621359}{5} \right)$ $\approx 1.0355987055 (\text{positif covid})$

Tahap 6 : Setelah centroid baru ditemukan, lakukan lagi langkah untuk menghitung euclidean distance dengan

centroid yang baru sampai penentuan *cluster* menjadi tetap sehingga didapat *cluster* tetapnya. ,

Tabel 16. Hasil clustering

No	Kecamatan	Cluster
1	Andir	1
2	Antapani	2
3	Arcamanik	1
4	Astanaanyar	3
.	.	.
30	Ujung Berung	2

E. Rata-Rata Variable

Pada tahapan ini akan menentukan karakteristik *cluster* seperti apa dengan menghitung mean dari variabel sehingga didapatkan perbandingan setiap *cluster*.

Tahap 1 : Menghitung rata-rata variabel untuk variabel obesitas yang termasuk kepada *cluster 1, cluster 2* dan *cluster 3*.

Tabel 17. Hitung rata-rata variable obesitas

Perhitungan	Proses Perhitungan
X	$\frac{1}{11}$
$= \frac{1}{11} \sum_{i=1}^n 1_i$	$554+177+1543+357+1766+1886+562+0+28+144$
Nilai sample 1	$2+201) \approx 774$
X	$\frac{1}{8}$
$= \frac{1}{8} \sum_{i=1}^n 1_i$	$\frac{1}{8}(279+0+975+620+8600+89+238+960) \approx 1470$
Nilai sample 2	
X	$\frac{1}{11}$
$= \frac{1}{11} \sum_{i=1}^n 1_i$	$\frac{1}{11}(1931+0+470+0+217+132+251+418+0+1039+527) \approx 453$
Nilai sample 3	

Berdasarkan perhitungan yang dihasilkan dari masing-masing variabel, didapatkan karakteristik pada masing-masing *cluster*.

Tabel 18. Hasil hitung rata-rata variable

cluster	obesitas	diabetes	Tubercu- losis	diare	hipertensi	Jumlah Penduduk	Positif covid
1	774	1092	58	2189	5217	112699	242
2	1470	750	28.5	2132	3031	83036	198
3	453	452	20	1094	2544	53308	125

Tahap 2 : menentukan karakteristik *cluster* dari rata-rata variabel yang didapat dengan cara membandingkan jumlah rata-rata variabel.

1. *Cluster 1* , memiliki karakteristik rata-rata variable diabetes , tuberculosis,diare,hipertensi, jumlah penduduk dan angka positif covid yang lebih tinggi dari *cluster 2* dan *3* .

2. *Cluster 2* ,memiliki karakteristik rata-rata variabel,obesitas yang lebih tinggi dari *cluster 1* dan 3.
3. *Cluster 3*,memiliki karakteristik rata-rata variable obesitas ,diabetes,tuberculosis,diare,hipertensi,jumlah penduduk dan positif *cov* lebih kecil dari *cluster 2* dan 1.

4. HASIL DAN PEMBAHASAN

A. Evaluasi Hasil /Pengujian Cluster

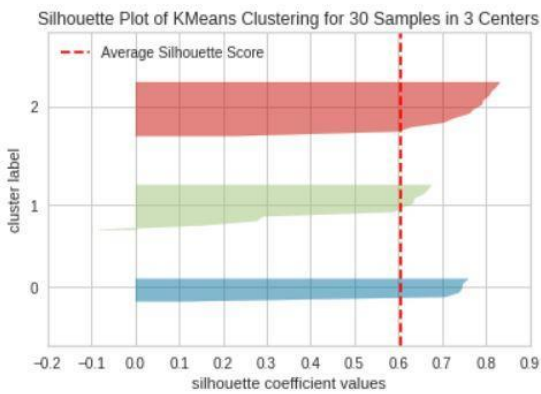
1. *Silhouette Coefficient*

Berikut adalah tabel proses pengujian terhadap data komorbid,populasi dan *covid* dengan memasukkan sebanyak 3 *cluster* yang berbeda dapat dilihat pada Tabel 4.4.

Table 19. Silhouette coefficient score

Jumlah Cluster	Silhouette Coefficient Score
1	0.7>
2	0.6>
3	0.8>

Dari hasil pengujian 3 *cluster* yang dimasukkan didapatkan *cluster* dengan jumlah 3 (tiga) yang memiliki nilai silhouette coefficient paling mendekati dengan nilai Si = 1 yaitu dengan nilai 0.8> pada jumlah *cluster* 3.



Gambar 2. Grafik silhouette

Nilai silhouette dapat dikatakan baik apabila bernilai positif karena titik pusat sudah berada di dalam *cluster* yang tepat sedangkan jika nilai silhouette dikatakan negatif menandakan terjadinya overlapping sehingga titik pusat berada diantara dua *cluster*.

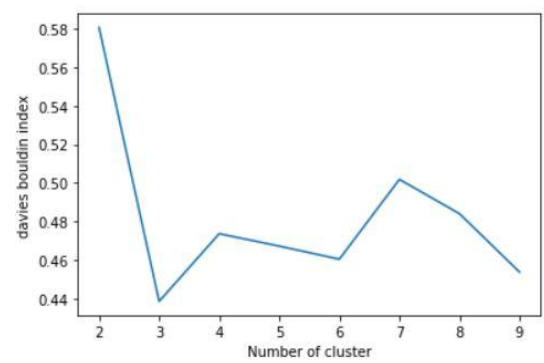
2. *Davies-bouldin index*

Pengujian dilakukan terhadap semua data sebanyak 30 data dan 7 variabel dengan memasukkan beberapa titik pusat secara acak dan jumlah *cluster* sebanyak 9 kemudian dilakukan perhitungan davies bouldin index yang hasilnya didapat pada tabel

Tabel 20. Davies-Bouldin index score

Jumlah Cluster	Davies-Bouldin index Score
2	0.580698159017849,
3	0.4384581806318664,
4	0.4736100293004229,
5	0.46715526070416385,
6	0.46036203228101585,
7	0.5018038950737858,
8	0.4839960637214656,
9	0.45368603694816023

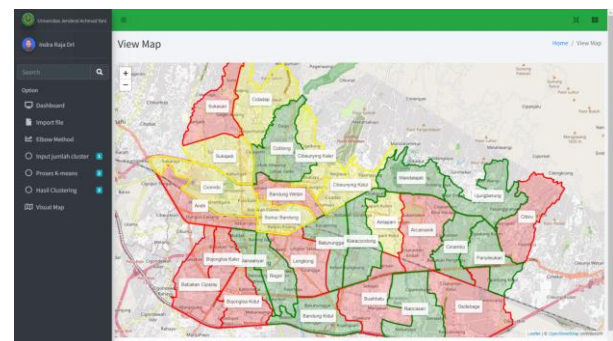
dari hasil yang didapat dengan melakukan validasi jumlah maksimal *cluster* ,jumlah *cluster* yang baik adalah 3 karena index yang paling rendah dari *cluster* yang lain dapat dilihat dalam grafik.



Gambar 3. Grafik davies-bouldin index

B. Visualisasi Cluster

Visualisasi *cluster* ini ditampilkan dengan menggunakan sistem informasi geografis untuk memetakan wilayah kecamatan kota bandung .terdapat 3 *cluster* yang dimana memiliki warna berbeda, untuk *cluster 1* berwarna merah dengan tingkat kasus *Covid-19* tinggi dan faktor pengaruh terhadap *Covid-19* juga tinggi ,*cluster 2* berwarna kuning dengan tingkat kasus yang lebih rendah dari *cluster 1* dan untuk *cluster 3* memiliki tingkatan kasus *Covid-19* dan faktor yang mempengaruhi lebih rendah dari *cluster 1* dan *cluster 3*.



Gambar 4. Clustering

5. KESIMPULAN

dapat disimpulkan bahwa wilayah dibagi berdasarkan karakteristik *clusternya* ,setiap *cluster* memiliki karakteristik yang berbeda,untuk wilayah yang termasuk kepada *cluster 1* yaitu andir,arcamanik

babakan ciparay, bandung kulon, batununggal, bojongloa kaler bojongloa kidul, cibeunying kidul, coblong, kiaracandong dan sukajadi memiliki karakteristik nilai rata-rata variable diabetes, tuberculosis, diare, hipertensi, jumlah penduduk dan angka positif covid yang lebih tinggi dari cluster 2 dan 3. untuk wilayah yang termasuk kepada cluster 2 antapani, buah batu, cibeunying kaler, cibiru, cicendo, regol, sukasari dan ujung berung memiliki karakteristik rata-rata variabel, obesitas yang lebih tinggi dari cluster 1 dan 3. untuk wilayah yang termasuk kepada cluster 3 yaitu Astana Anyar, bandung kidul, bandung wetan, cidadap, cinambo, gedebage, lengkong, mandalajati, panyileukan, rancasari dan sumur bandung memiliki karakteristik nilai rata-rata variabel obesitas, diabetes, tuberculosis, diare, hipertensi, jumlah penduduk dan positif covid lebih kecil dari cluster 2 dan

DAFTAR PUSTAKA

- [1] R. Bag, M. Ghosh, B. Biswas, and M. Chatterjee, "Understanding the spatio-temporal pattern of COVID-19 outbreak in India using GIS and India's response in managing the pandemic," *Regional Science Policy and Practice*, vol. 12, no. 6, pp. 1063–1103, Dec. 2020, doi: 10.1111/rsp3.12359.
- [2] K. Tantrakarnapa, B. Bhopdhornangkul, and K. Nakhaapakorn, "Influencing factors of COVID-19 spreading: a case study of Thailand," *Journal of Public Health (Germany)*, 2020, doi: 10.1007/s10389-020-01329-5.
- [3] C. Sohrabi *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76. Elsevier Ltd, pp. 71–76, Apr. 01, 2020. doi: 10.1016/j.ijssu.2020.02.034.
- [4] A. Mollalo, B. Vahedi, and K. M. Rivera, "GIS-based spatial modeling of COVID-19 incidence rate in the continental United States," *Science of the Total Environment*, vol. 728, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138884.
- [5] M. Yıldırım and A. Güler, "COVID-19 severity, self-efficacy, knowledge, preventive behaviors, and mental health in Turkey," *Death Studies*, pp. 1–8, 2020, doi: 10.1080/07481187.2020.1793434.
- [6] B. Gavin, J. Lyne, and F. McNicholas, "Mental health and the COVID-19 pandemic," *Irish Journal of Psychological Medicine*, vol. 37, no. 3. Cambridge University Press, pp. 156–158, Sep. 01, 2020. doi: 10.1017/ipm.2020.72.
- [7] Y. Gao, M. Liu, Y. Chen, S. Shi, J. Geng, and J. Tian, "Association between tuberculosis and COVID-19 severity and mortality: A rapid systematic review and meta-analysis," *Journal of Medical Virology*, vol. 93, no. 1. John Wiley and Sons Inc, pp. 194–196, Jan. 01, 2021. doi: 10.1002/jmv.26311.
- [8] I. Huang, M. A. Lim, and R. Pranata, "Diabetes mellitus is associated with increased mortality and severity of disease in COVID-19 pneumonia – A systematic review, meta-analysis, and meta-regression: Diabetes and COVID-19," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 4, pp. 395–403, Jul. 2020, doi: 10.1016/j.dsx.2020.04.018.
- [9] Z. Kang *et al.*, "Obesity is a potential risk factor contributing to clinical manifestations of COVID-19," *International Journal of Obesity*, vol. 44, no. 12, pp. 2479–2485, Dec. 2020, doi: 10.1038/s41366-020-00677-2.
- [10] X. Yang, J. Zhao, Q. Yan, S. Zhang, Y. Wang, and Y. Li, "A case of COVID-19 patient with the diarrhea as initial symptom and literature review," *Clinics and Research in Hepatology and Gastroenterology*, vol. 44, no. 5, pp. e109–e112, Oct. 2020, doi: 10.1016/j.clinre.2020.03.013.
- [11] C. H. Zhang and G. G. Schwartz, "Spatial Disparities in Coronavirus Incidence and Mortality in the United States: An Ecological Analysis as of May 2020," *Journal of Rural Health*, vol. 36, no. 3, pp. 433–445, Jun. 2020, doi: 10.1111/jrh.12476.
- [12] covid19.go.id, "satgas penting memahami faktor penyebab lonjakan agar terhindar gelombang baru Covid-19," Oct. 2021, Accessed: Feb. 01, 2022. [Online]. Available: <https://covid19.go.id/artikel/2021/10/05/satgas-penting-memahami-faktor-penyebab-lonjakan-agar-terhindar-gelombang-baru-Covid-19>
- [13] R. S. Wardani, Purwanto, Sayono, and A. Paramananda, "Clustering tuberculosis in children using K-Means based on geographic information system," in *AIP Conference Proceedings*, Jun. 2019, vol. 2114. doi: 10.1063/1.5112483.
- [14] Z. Kang *et al.*, "Obesity is a potential risk factor contributing to clinical manifestations of COVID-19," *International Journal of Obesity*, vol. 44, no. 12, pp. 2479–2485, Dec. 2020, doi: 10.1038/s41366-020-00677-2.
- [15] A. Mollalo, B. Vahedi, and K. M. Rivera, "GIS-based spatial modeling of COVID-19 incidence rate in the continental United States," *Science of the Total Environment*, vol. 728, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138884.
- [16] D. Marutho, S. Hendra Handaka, and E. Wijaya, "The Determination of Cluster Number at k-mean using Elbow Method and Purity Evaluation on Headline News."
- [17] statmat.net, "Cara Menghitung Nilai Rata-rata dalam Statistika dan Contohnya," *statmat.net*, 2020. <https://www.statmat.net/cara-menghitung-nilai-rata-rata/> (accessed Feb. 01, 2022).
- [18] F. Nur, R. Fauzan, J. Aziz, B. Darma Setiawan, and I. Arwani, "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [19] M. D. Kartikasari, "Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan," *Jambura Journal of Mathematics*, vol. 3, no. 2, pp. 187–196, Jul. 2021, doi: 10.34312/jjom.v3i2.10942.

- [20] A. Murad and B. F. Khashoggi, "Using GIS for disease mapping and *clustering* in Jeddah, Saudi Arabia," *ISPRS International Journal of Geo-Information*, vol. 9, no. 5, May 2020, doi: 10.3390/ijgi9050328.