

Penerapan Metode K-Means dalam Mengelompokkan Tingkat Kesembuhan Penderita Covid-19

IMPLEMENTATION OF K-MEANS METHOD IN GROUPING THE CURE RATE OF COVID-19 SUFFERERS

Ir. Hastha Sunardi^{*1}, Fery Anthony²

^{1,2} Universitas IGM : Jl. Jend. Sudirman km. 4,5, Palembang 30129, Indonesia

¹ Program Studi Sistem Komputer Universitas IGM – Palembang

e-mail: ^{*1}hastha_s@uigm.ac.id, ²feryanthony@uigm.ac.id

Abstrak

Pandemi Virus Corona (Covic-19) yang mulai terdengar di Indonesia sekitar akhir Februari 2020 seakan terus berlanjut, dimana tak satupun pakar dari satu disiplin ilmu yang mampu memprediksi pandemi ini, kapan berakhir. Berbeda dengan penelitian sebelumnya yang lebih fokus pada tingkat penyebaran Covid-19. Penelitian ini menerapkan metode K-Means menitik beratkan pada pengelompokan tingkat kesembuhan penderita terpapar Covid-19 dengan proses penyembuhan isolasi mandiri. Tidak adanya perbedaan jenis kelamin dalam proses penularan, sehingga data gender digunakan sebagai id_penderita dan bukan berjenis nominal, sehingga tidak perlu proses inisialisasi. Oleh karena itu, cukup menarik untuk dianalisis bahwa usia, lama isolasi, dan jumlah anggota keluarga kaitannya dengan penyebaran Covid-19 dan lebih difokuskan pada pengelompokan tingkat penyembuhan dengan isolasi mandiri. Adapun nilai rerata untuk setiap cluster, diperoleh nilai rerata tertinggi untuk usia terdapat pada cluster 3 sebesar 53,8 tahun dan terendah pada cluster 2 sebesar 30 tahun. Untuk lama isolasi nilai rerata terlama pada cluster 1 sebesar 29,7 hari dan tercepat pada cluster 3 yaitu 14 hari.

Kata kunci — Covid-19, isolasi mandiri, cluster dan tingkat penyembuhan

Abstract

The coronavirus pandemic (Covic-19) that began to be heard in Indonesia around the end of February 2020 seems to continue, where none of the experts from one discipline are able to predict this pandemic, when it ends. Unlike previous research that focused more on the extent of the spread of Covid-19. This research applies the K-Means method to focus on grouping the cure rate of patients exposed to Covid-19 with a self-isolation healing process. There is no difference in sex in the transmission process, so gender data is used as a id_penderita and not nominal type, so there is no need for initialization process. Therefore, it is quite interesting to analyze that the age, length of isolation, and number of family members is related to the spread of Covid-19 and is more focused on grouping healing rates with self-isolation. As for the average value for each cluster, the highest average value for age is found in cluster 3 at 53.8 years and the lowest in cluster 2 by 30 years. For the length of isolation the longest average value in cluster 1 is 29.7 days and the fastest in cluster 3 is 14 days.

Kata kunci — Covid-19, self-isolation, clusters and healing rates

1. PENDAHULUAN

Pandemi Virus Corona (Covic-19) yang mulai terdengar di Indonesia sekitar akhir Februari 2020 seakan terus berlanjut, dimana tak satupun pakar dari satu disiplin ilmu yang mampu memprediksi pandemi ini, kapan berakhir. Wabah penyakit yang menyerang area pernafasan ini, penyebarannya cukup mudah melalui *droplet* dari pasien terjangkit, sehingga penyebarannya sangat pesat [1]. Penyakit ini memiliki simptom bermula demam, batuk kering, sakit kepala, sakit tenggorokan, nyeri dan lelah disertai tidak ada

nafsu makan disebabkan oleh virus SARS-CoV-2 (*Severe acute respiratory syndrome coronavirus 2*), virus ini dapat juga mengakibatkan simptom mematikan seperti sesak napas, rasa sakit pada dada, dan kesulitan untuk berbicara dan bernapas, hingga dapat berujung kepada kematian. Orang dengan sistem imun yang lemah dan orang lanjut usia (lansia) terutama sangat rentan terhadap infeksi virus ini beserta dampak-dampaknya. Virus ini sangat cepat untuk menyebar karena metode penularan dari manusia ke manusia-nya yang sangat efektif yaitu melalui *close contact* (kontak dekat) dan *respiratory droplets* seperti batuk atau bersin. Virus ini ditentukan berasal mula dari sebuah pasar basah pada kota Wuhan di Cina, dan telah menyebar ke hampir seluruh negara di dunia. Faktanya sekarang Covid-19 membuktikan diri mampu menular antar manusia. Penularan sangat cepat hingga Organisasi Kesehatan Dunia WHO menetapkan Pandemi virus Corona atau COVID-19 pada tanggal 11/3/2020 [2].

Pada tanggal 2 Maret 2020 di Indonesia sebanyak dua kasus positif Covid-19 pertama kali dilaporkan. Terkonfirmasi berjumlah 1.528 kasus dan 136 kasus kematian pada 31 Maret 2020. Tingkat kematian *Covid-19* di Indonesia sebesar 8,9%, adalah jumlah tertinggi di Asia Tenggara [3].

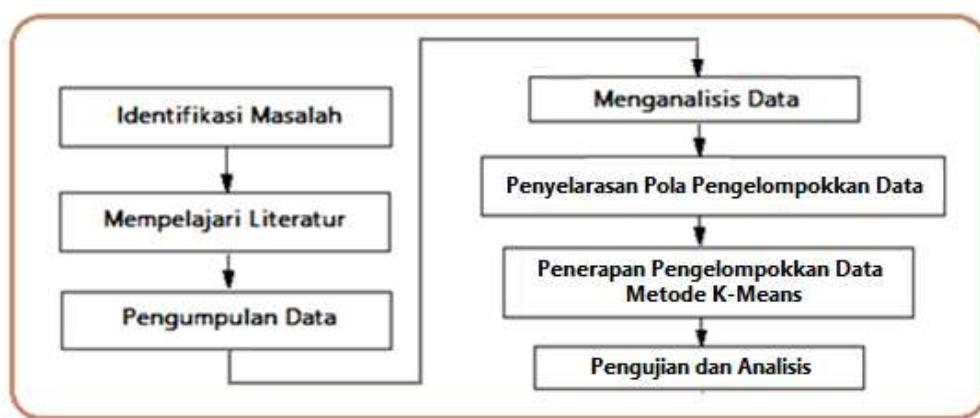
Begitu cepat penyebaran wabah ini, menarik beberapa peneliti. Beberapa penelitian memanfaatkan algoritma klasterisasi k-means untuk berbagai keperluan, antara lain pengelompokan jumlah kematian penderita covid-19 berdasarkan negara di benua asia.[4]. Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta [5], menentukan tingkat penyebaran pandemi covid-19 di Indonesia [6]. Untuk sebaran kasus Covid-19 di daerah Sumatera Barat, dengan menggunakan Metode K-Means, bagaimana pengelompokan data ke dalam beberapa kelompok dalam kasus penyebaran covid-19 [7].

Berbeda dengan penelitian lainnya yang lebih memfokus diri pada tingkat penyebaran Covid-19. Penelitian ini menitik beratkan pada pengelompokan tingkat kesembuhan penderita terpapar Covid-19 dengan proses penyembuhan dijalani di rumah atau yang dikenal dengan nama Isolasi Mandiri. Tingkat penyembuhan dengan cara isolasi mandiri, sebenarnya juga terkandung makna tingkat penyebaran terdekat, yaitu dalam satu keluarga. Bagaimana kita mendapatkan gambaran hubungan perbandingan jumlah anggota keluarga dan jumlah penderita dari satu keluarga tersebut terhadap tingkat penyebaran maupun tingkat kesembuhannya. Namun dalam penelitian ini, difokuskan pada tingkat penyembuhan penderita covid-19 dengan cara isolasi mandiri.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Dalam metode ini cara untuk memperoleh dan mengumpulkan data-data ilmiah yang dibutuhkan dengan fungsi dan tujuan tertentu. Metode penelitian dapat dilihat pada gambar di bawah ini:



Gambar 1. Diagram Kerangka Kerja

2.2 Objek Penelitian

Adapun yang menjadi objek dalam penelitian kuantitatif ini, yaitu: (1). Objek penelitian korelasi, dimana dilakukan analisis korelasi untuk melihat keterkaitan antar variabel. Pada penelitian ini mengambil, (2). Objek penelitian komparatif, yakni melakukan perbandingan atau komparatif Usia, Lama Isolasi dan Jumlah Anggota Keluarga. Serta melakukan perbandingan beberapa metode dalam menentukan jumlah kluster yang optimal untuk diterapkan dalam Klasifikasi K-Means.

2.3 Algoritma K-Means

K-Means merupakan salah satu metode pengelompokan data non-hirarki yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data berkarakteristik sama dimasukkan ke dalam satu cluster yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam cluster yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster [8]. K-Means menggunakan “mean” yaitu rata-rata sebagai pusat cluster-nya, sehingga K-Means tidak robust atau tidak cukup baik terhadap data yang memiliki outlier. Sehingga, untuk mengatasi hal tersebut dapat digunakan K-Medoids untuk mengelompokkan data yang memiliki outlier [9].

K-Means merupakan *algoritma clustering* yang berulang-ulang. *Algoritma K-Means* menetapkan nilai-nilai *cluster* (*k*) secara *random*, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, *mean* atau “*means*”. Kemudian menghitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus *Euclidian* hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasisifkasikan setiap data berdasarkan kedekatannya dengan *centroid*. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil). [10]. Berikut adalah tahapan untuk melakukan optimasi menggunakan *algoritma K-Mean* :

- 1) Tetapkan pusat cluster secara acak
- 2) Hitunglah dari setiap data yang ada, jarak terhadap setiap cluster.

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2-1)$$

Keterangan:

(x,y) = Koordinat objek
(s,t) = Koordinat *Centroid*
i = Banyaknya objek

- 3) Anggota baru dari suatu cluster, diperoleh jika memiliki jarak terkecil dari pusat clusternya dari suatu data.
- 4) Hitung pusat cluster baru, dengan rumus berikut :

$$Vij = \frac{1}{Ni} \sum_{k=0}^{Ni} Xkj \quad (2-2)$$

Keterangan:

Vij = *Centroid* rata – rata pada *Cluster* ke – i untuk variabel ke – j
Ni = Jumlah anggota *Cluster* ke – i
i,k = Indeks dari *Cluster*

j = Indeks variabel

X_{kj} = Nilai data ke-k variabel ke-j untuk *Cluster* tersebut

- 5) Kembali ke langkah 2 hingga posisi data sudah tidak mengalami perubahan.

3. HASIL DAN PEMBAHASAN

3.1 Evaluasi dan Visualisasi Data

Pada langkah pertama dari penelitian ini, dilakukan eksplorasi data. Dataset Penderita terdiri atas ID Penderita, Usia, Lama Isolasi dan Jumlah Anggota Keluarga. Selanjutnya untuk proses perhitungan menggunakan Pemrograman Studio R dilakukan import data data dari excel, lalu kita dapat melihat data atau struktur data, dan data siap untuk diolah. Hal ini dilakukan untuk mendapatkan pengetahuan dasar mengenai data dan mungkin diperlukan dalam menemukan atau menentukan assumsi awal.

Struktur data gender (jenis kelamin) adalah faktor (kualitatif) sehingga untuk menganalisis atribut ini lebih tepat dengan melihat frekuensi atau persentasi antar faktor (dalam hal ini pria dan wanita). Namun dalam penelitian ini dipandang bahwa kasus pengelompokkan lingkungan terhadap penularan covid-19, tidak ada perbedaan jenis kelamin dalam proses penularan, sehingga data gender digunakan sebagai id_penderita dan bukan berjenis nominal, sehingga tidak perlu proses inisialisasi. Oleh karena itu, cukup menarik untuk dianalisis bahwa usia (*usia*), lama isolasi (*lama_isolasi*), dan jumlah anggota keluarga (*jl_angkel*) kaitannya dengan penyebaran covid-19 dan lebih difokuskan pada pengelompokan tingkat penyembuhan dengan isolasi mandiri. Atribut-atribut yang gunakan dalam pengelompokan. Pertimbangannya, karena siklus penularan covid-19 bukan dari udara, namun dari lingkungan terdekat (*droplet*) seseorang sehingga potensi penularan dalam satu keluarga cukup potensial. Visualisasi data terlihat pada Tabel 1.

Tabel 1. Data Penderita

	id_penderita	usia	lama_isolasi	jl_angkel
1	L1	58	14	6
2	P1	55	14	6
3	L2	46	30	5
4	L3	35	15	5
5	P2	34	15	5
6	P3	57	13	4
..
14	P7	67	13	2
15	P8	24	14	2
16	L8	55	29	4
17	L9	45	30	2
18	P9	50	14	5
19	L10	50	10	3

Showing 1 to 19 of 100 entries 4 total columns

3.2 Deskripsi Data

Pada dataset, dapat dideskripsikan bahwa atribut id-penderita berisikan data berupa simbol jenis kelamin laki-laki (L) atau perempuan (P) disertai nomor urutnya. Sedangkan untuk usia tertinggi 67 tahun dan terendah 23 tahun, lama isolasi tertinggi 30 hari dan tercepat 10 hari serta jumlah anggota keluarga terbanyak 6 orang dan terkecil 2 orang. Tahapan evaluasi data harus dilakukan sebelum proses perhitungan pengelompokan (*cluster*). Hasil evaluasi tidak ditemukan nilai kosong (*null*) atau *missing data*. Hal ini dapat ditunjukkan pada deskripsi data pada Gambar 1. Dengan tanpa nilai kosong, maka tahapan pengelompokan berikutnya adalah proses transformasi data.

```
> # Lihat data, apakah lengkap terisi?  
> summary(RisetCovid19)  
  
id_penderita      usia      lama_isolasi    jl_angkel  
Length:100        Min.   :23.00   Min.   :10.00   Min.   :2.00  
Class :character  1st Qu.:34.00   1st Qu.:13.75  1st Qu.:3.00  
Mode  :character  Median :45.00   Median :14.00   Median :4.00  
                  Mean   :45.07   Mean   :16.21   Mean   :3.79  
                  3rd Qu.:55.00   3rd Qu.:15.00  3rd Qu.:4.00  
                  Max.   :67.00   Max.   :30.00   Max.   :6.00  
  
Data Lengkap !
```

Gambar 2. Deskripsi Data

3.3 Transformasi Data

Transformasi data ini dilakukan, dengan tujuan utama untuk mengubah skala pengukuran data asli menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari analisis, dengan proses normalisasi. Normalisasi adalah proses pengelompokan atribut ke dalam hubungan yang terstruktur dengan baik dan bebas dari anomali (Lee, 1995). Normalisasi digunakan untuk mentransformasi sebuah atribut neterik diskalakan dalam jangkauan yang lebih kecil seperti -1,0 sampai 1,0. Teknik yang digunakan untuk operasi ini adalah normalisasi min-max, normalisasi Z-Score dan normalisasi skala desimal.

Normalisasi skala desimal, proses normalisasi diperoleh dengan melakukan pergeseran titik decimal dari value sebuah atribut. Jumlah titik decimal yang digeser tergantung dari nilai absolut maksimum dari atribut tersebut.

Pada penelitian ini menggunakan cara skala. Pada sampel data yang ditunjukkan pada Tabel 1, atribut Usia (*usia*), lama isolasi (*lama_isolasi*) dan Jumlah Anggota Keluarga (*jl_angkel*) bernilai numerik, namun nilai jangkauan umur, lama isolasi dan jumlah anggota keluarga berbeda, yakni jangkauan umur maksimal 100 tahun, lama isolasi maksimal satu bulan (30 hari) dan jumlah anggota keluarga maksimum 10 orang, sehingga perlu dilakukan transformasi data menggunakan penskalaan. Melalui Pemrograman Studio R, diperoleh hasil sebagaimana ditunjukkan pada Tabel 2.

3.4 Analisis Cluster

Analisis cluster bertujuan untuk mengelompokkan objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam cluster yang sama. Berbeda dengan analisis multivariat yang lainnya analisis cluster tidak mengestimasi himpunan variabel secara empiris sebaliknya menggunakan himpunan variabel yang

ditentukan untuk mengelompokkan objek. Himpunan variabel cluster adalah suatu himpunan variabel yang merepresentasikan karakteristik yang dipakai untuk mengelompokkan objek yang diteliti. Ciri-ciri analisis cluster adalah homogenitas (kesamaan) yang tinggi antar anggota dalam satu cluster dan heterogenitas (perbedaan) yang tinggi antar cluster yang satu dengan cluster yang lainnya [3]. Solusi analisis cluster bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Analisis cluster pada metode non-hirarki mengelompokkan objek dengan menentukan bakal cluster terlebih dahulu dan banyaknya cluster sudah ditentukan sebelumnya. Sedangkan metode hirarki banyaknya cluster tidak ditentukan.

Tabel 2 Nilai Hasil Transformasi Data

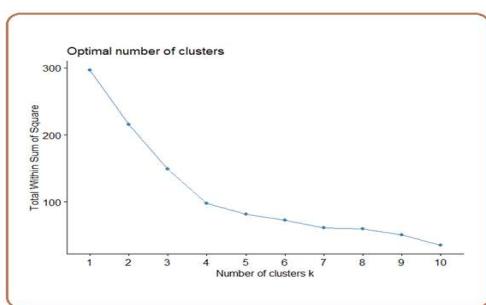
	usia	lama_isolasi	jl_angkel
1	0.959807768	-0.41214251	2.1924640
2	0.737114550	-0.41214251	2.1924640
3	0.069034897	2.57169466	1.2003988
4	-0.747506900	-0.22565269	1.2003988
5	-0.821737973	-0.22565269	1.2003988
6	0.885576695	-0.59863233	0.2083337
7	-0.821737973	2.57169466	0.2083337
8	-0.747506900	2.57169466	0.2083337
9	-0.821737973	2.38520484	0.2083337
10	-0.005196175	2.19871502	1.2003988
11	1.479425276	-0.41214251	-0.7837315

Showing 1 to 11 of 100 entries, 3 total columns

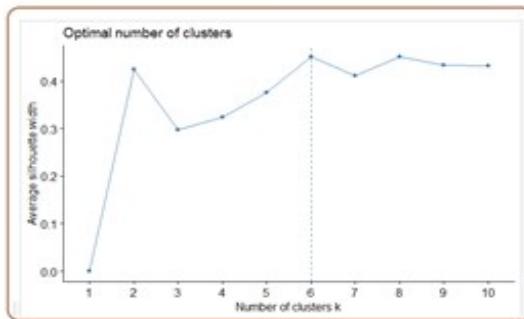
3.5 Menentukan Jumlah Cluster

Algoritma *K-Means* merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat cluster awal. Pada tahap ini pusat cluster dipilih secara acak dari sekumpulan populasi data [12].

Selanjutnya akan mencari nilai k optimum, karena analisis k-means merupakan analisis non hirarki, maka nilai k ditentukan sendiri oleh peneliti, disini peneliti menggunakan nilai k=3, namun untuk membuktikan nilai k yang digunakan paling optimal bisa dengan membandingkan tiga metode yaitu metode wss, metode sillhouette, dan metode gap_statistic. Melalui Pemrograman Studio R, diperoleh keluaran berupa grafik yang menunjukkan jumlah cluster k yang ditunjukkan pada Gambar 3 dan Gambar 4. Dari kedua metode yang digunakan (Metode Elbow dan Metode Silhouette), menghasilkan nilai k yang berbeda, yaitu k = 4 dan k = 6, sebagaimana ditunjukkan Gambar 4 dan Gammbar 5 diatas. Untuk k = 4, menghasilkan 2 cluster dari 4 cluster yang saling bersinggungan. Sedangkan untuk k = 6, justru menghasilkan 2 cluster yang saling beririsan.



Gambar 3. Optimum nilai k Elbow



Gambar 4. Optimum nilai k Silhouette

3.6 Visualisasi Hasil Cluster

Hasil cluster divisualisasikan dalam tiga bentuk, yaitu bentuk vektor, grafik dan tabel. Dengan menggunakan $k=3$, pada Gambar 5, diketahui cluster 1 memiliki jumlah data sebesar 12, cluster 2 memiliki jumlah data sebesar 30, dan cluster 3 memiliki jumlah data sebesar 58. Setelah diketahui bahwa data 1–100 masuk ke cluster mana, selanjutnya didapatkan nilai within cluster sum of squares dari cluster k-means dengan $k=3$ yaitu untuk cluster 1 sebesar 15.74678, cluster 2 sebesar 30.61118, dan cluster 3 sebesar 100.36655. Dengan nilai between_SS atau between sum of squares dibagi nilai total_SS sebesar 50.6%.

```
> final=kmeans(covid19fix,3)
> print(final)
K-means clustering with 3 clusters of sizes 12, 30, 58

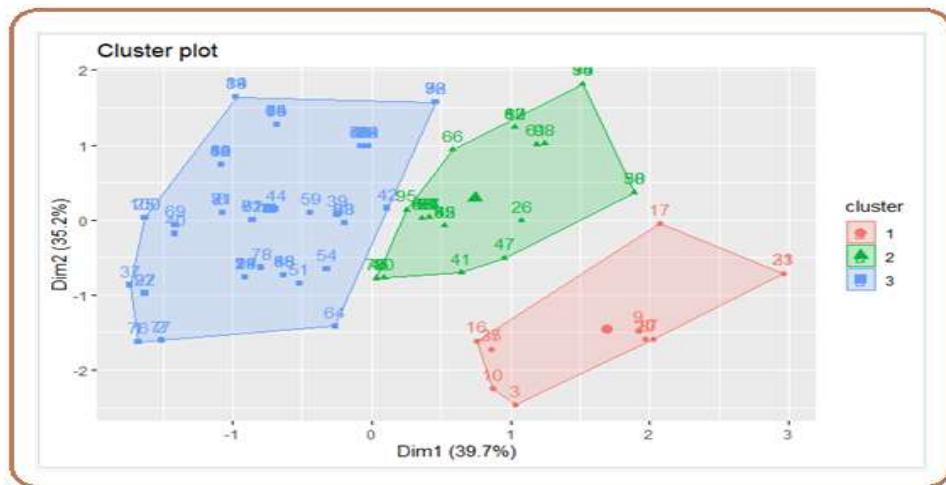
Cluster means:
  usia lama_isolasi  jl_angkel
1 -0.3454219  2.5095314  0.04298949
2 -1.1161879  -0.2070037 -0.08928586
3  0.6488052  -0.4121425  0.03728797

Clustering vector:
 [1] 3 3 1 2 2 3 1 1 1 3 2 3 3 2 1 1 3 3 1 2 3 1 3 3 2 1 3
[29] 3 3 1 3 3 2 1 3 3 2 3 3 2 3 3 3 2 2 2 3 3 2 3 3 3 3 2 2 2
[57] 2 3 3 2 2 3 3 3 3 2 2 3 3 3 3 3 3 3 3 3 3 3 2 2 3 2 2 2
[85] 3 3 2 3 3 2 3 3 3 2 2 3 2 3 3 3 2 2 3 2 3 3 3 2 2 3 2 2 2

Within cluster sum of squares by cluster:
[1] 15.74678 30.61118 100.36655
  (between_SS / total_SS =  50.6 %)

Available components:
[1] "cluster"      "centers"       "totss"
[4] "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"          "ifault"
```

Gambar 5 Hasil Cluster Bentuk Vektor



Gambar 6. Hasil Cluster Bentuk Grafik

Agar lebih jelas terlihat data mana yang termasuk pada cluster 1, cluster 2 atau cluster 3, terlihat pada Tabel 3 berikut :

Tabel 3. Hasil Cluster Bentuk Tabel

	usia	lama_isolasi	jl_angkel	final.cluster
1	0.959807768	-0.41214251	2.1924640	3
2	0.737114550	-0.41214251	2.1924640	3
3	0.069034897	2.57169466	1.2003988	1
4	-0.747506900	-0.22565269	1.2003988	2
5	-0.821737973	-0.22565269	1.2003988	2
6	0.885576695	-0.59863233	0.2083337	3
7	-0.821737973	2.57169466	0.2083337	1
8	-0.747506900	2.57169466	0.2083337	1
9	-0.821737973	2.38520484	0.2083337	1
10	-0.005196175	2.19871502	1.2003988	1
11	1.479425276	-0.41214251	-0.7837315	3
12	-1.638279771	-0.78512216	-0.7837315	2
13	-0.079427248	-0.59863233	-0.7837315	3
14	1.627887421	-0.59863233	-1.7757966	3
15	-1.564048698	-0.41214251	-1.7757966	2
16	0.737114550	2.38520484	0.2083337	1
17	-0.005196175	2.57169466	-1.7757966	1
18	0.365959188	-0.41214251	1.2003988	3
19	0.365959188	-1.15810180	-0.7837315	3

Showing 1 to 20 of 100 entries, 4 total columns

```
+ # Melihat nilai rerata pada setiap cluster
+ # Lakukan analisis !
+ summarise_all("mean")
# A tibble: 3 x 4
  cluster usia lama_isolasi jl_angkel
* <int> <dbl>      <dbl>     <dbl>
1       1  40.4        29.7     3.83
2       2  30.0        15.1     3.7
3       3  53.8        14.0     3.83
```

Gambar 7. Nilai Rerata Per Cluster

3.7 Analisis Hasil Clustering

Visualisasi cluster bentuk vector, dapat dilihat masing-masing data masuk dalam cluster 1, 2 atau 3. Untuk dapat melihatnya, dapat ditunjukkan pada Tabel 3. Terlihat data 1 dan data 2 masuk ke cluster 3, data 3 masuk ke cluster 1, data 4 & data 5 masuk cluster 3. Demikian halnya dengan data lainnya hingga data ke 100 masuk ke cluster mana dengan melihat clustering vectornya. Dimana dari hasilnya cluster, cluster 3 memiliki data terbanyak sebanyak 58 dan cluster 1 memiliki data terkecil sebanyak 12.

Sedangkan visualisasi cluster berupa grafik, dapat terlihat pada Gambar 4, dimana terlihat jelas, setiap cluster memiliki warna yang berbeda. Cluster 1 berwarna merah dengan data sedikit diantaranya 17, 23 dan lainnya, sehingga luasnya juga kecil. Cluster 2 berwarna hijau dengan data diantaranya 26, 89 dan lainnya. Sedangkan cluster 3 berwarna biru dengan memiliki anggota terbanyak, tampak bagannya luas dan dengan data 69, 88 dan lainnya. Visualisasi cluster dengan tabel, nampak jelas sekali terlihat masing-masing data dari 1 – 100 masuk dalam cluster 1, cluster 2 atau cluster 3.

Ketiga bentuk visualisasi tampaklah jelas bagimana penderita covid-19 dengan data penderita berupa atribut usia, lama isolasi, jumlah anggota keluarga dan lingkungan terkelompokkan dalam 3 cluster. Adapun nilai rerata untuk setiap cluster pada Gambar 7, terlihat bahwa nilai rerata tertinggi untuk usia terdapat pada cluster 3 sebesar 53,8 tahun dan terendah pada cluster 2 sebesar 30 tahun. Untuk Lama isolasi nilai rerata terlama pada cluster 1 sebesar 29,7 hari dan tercepat pada cluster 3 yaitu 14 hari.

4. KESIMPULAN

Berdasarkan hasil analisis dari hasil pengukuran kompresi dari berbagai posisi citra beraturan da citra tak beraturan, dapat disimpulkan :

1. Pemilihan jumlah pengelompokan (Cluster) untuk oaring terpapar covid-19 dengan atribut usia, lama isolasi (isolasi mandiri) dan jumlah anggota keluarga mengahsilkan pengelompokan ideal (tidak bersinggungan dan beririsan).
2. Pada setiap cluster memiliki warna yang berbeda. Cluster 1 berwarna merah dengan data sedikit diantaranya 17, 23 dan lainnya, sehingga luasnya juga kecil. Cluster 2 berwarna hijau dengan data diantaranya 26, 89 dan lainnya. Sedangkan cluster 3 berwarna biru dengan memiliki anggota terbanyak, tampak bagannya luas dan dengan data 69, 88 dan lainnya. Visualisasi cluster dengan tabel, nampak jelas sekali terlihat masing-masing data dari 1 – 100 masuk dalam cluster 1, cluster 2 atau cluster 3.
3. Adapun nilai rerata untuk setiap cluster pada Gambar 7, terlihat bahwa nilai rerata tertinggi untuk usia terdapat pada cluster 3 sebesar 53,8 tahun dan terendah pada cluster 2 sebesar 30 tahun. Untuk Lama isolasi nilai rerata terlama pada cluster 1 sebesar 29,7 hari dan tercepat pada cluster 3 yaitu 14 hari.

UCAPAN TERIMA KASIH

Terima kasih kepada Dekan Fasilkom Universitas IGM Bapak Dr. Juhaini Alie, M.M., dan Kepala LLPM Universitas IGM Ibu Tertiaavini, M.Kom yang telah memberikan dukungan dan motivasi selama tahapan penelitian sampai dituangkan dalam bentuk artikel.

DAFTAR PUSTAKA

- [1] Meida Cahyo Untoro , Leslie Anggraini , Maria Andin , Hesti Retnosar and M. Anas Nasrulloh, "Penerapan metode *k-means clustering* data COVID-19 di Provinsi Jakarta," *Jurnal Ilmiah Sistem Informasi*, vol. 11, no. 2, pp. 59–68, 2021
- [2] R. A. Indraputra and R Fitriana, "K-Means Clustering Data Covid-19," *Jurnal Teknik Industri*, vol. 10, no. 3, pp. 275–282, 2020.
- [3] A.R. Setiawan, "Lembar Kegiatan Literasi Saintifik untuk Pembelajaran Jarak Jauh Topik Penyakit Coronavirus 2019 (COVID-19)," *Edukatif: Jurnal Ilmu Pendidikan* vol. 2, no.1, pp. 28-37, 2020.
- [4] Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita Covid-19 Berdasarkan Negara di Benua Asia," *Paradig. Informatika. dan Komputer*, vol. 22, no. 2, pp. 183–188, 2020.
- [5] A. Solichin and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) di DKI Jakarta Menggunakan K-Means," *Fountain Informatics J.*, vol. 5, no. 2, pp. 52-61, 2020.
- [6] P. P. P. A. N . F. I. R . Z. Nayuni Dwitri , Jose A Tampubolon , Sandi Prayoga and D. Hartama, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia," *JTI (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 128–132, 2020.
- [7] Supranto, "Analisis Multivariat: Arti dan Interpretasi," *Jakarta: PT. Rineka Cipta*, 2004.
- [8] Triyanto, W. A, "Algoritma K-Medoids untuk Penentuan Strategi Pemasaran," *Simetris*, pp. 183–188, 2015.
- [9] Larasati, A., "Metode K-Medoids pada Data dengan Penciran," *Tugas Akhir*, Yogyakarta: Universitas Gadjah Mada, 2017.
- [10] D. Permata Sari, "Implementasi Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 di Sumatera Barat, " *Computer Based Information System Journal*, vol. 09, no. 1, p. 50-56, 2021.
- [11] Heeseok Lee, "Justfying Database Normalization: A Cost/Benefit Model, Information Processing & Management," *Elvisevier Science Ltd.*, vol. 31, no.1, pp. 59-67, 1995.
- [12] Yulia Darmi and Agus Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk, " *Jurnal Media Infotama*, vol. 12, no. 2, pp. 148-157, 2016.