

Volume 3 Nomor 1 Maret 2020

e-ISSN 2614-4204

p-ISSN 2615-465X

***MAJA**math*

JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA



Diterbitkan :
Program Studi Pendidikan Matematika
Universitas Islam Majapahit



MAJAMATH: JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA ✓

[Google Scholar](#) [Website](#) [Editor URL](#)

[Prodi Pendidikan Matematika Universitas Islam Majapahit](#)

P-ISSN : 2615445X | E-ISSN : 26144204

S4 Accredited

Garuda Indexed

0,00 Impact	9 H5-index	288 Citations 5yr	302 Citations
----------------	---------------	----------------------	------------------

Previous **1** Next

Page 1 of 1 | Total Records 1

Get More with
SINTA Insight

Go to Insight



MAJAMATH: JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA

! [PRODI PENDIDIKAN MATEMATIKA UNIVERSITAS ISLAM MAJAPAHIT](#)" [P-ISSN : 2615465X](#) <> [E-ISSN : 26144204](#)# **0**
Impact

302
Google
Citations

Sinta 4
Current
Accreditation

[Google Scholar](#)
[Garuda](#)
[Website](#)
[Editor URL](#)

History Accreditation

2018 2019 2020 2021 2022 2023 2024 2025 2026

Garuda[Google Scholar](#)

Search...

[Simulasi Clustering Provinsi di Indonesia dalam Penyebaran Covid-19 Berdasarkan Indikator Kesehatan Masyarakat Menggunakan Algoritma Gaussian Mixture Model](#)

Prodi Pendidikan matematika Universitas Islam Majapahit (UNIM), Mojokerto, Indonesia
& [MAJAMATH: Jurnal Matematika dan Pendidikan Matematika Vol. 5 No. 1 \(2022\): Vol. 5 No. 1 Maret 2022 43 - 60](#)

 2022 [DOI: 10.36815/majamath.v5i1.1699](#) [Accred : Sinta 4](#)

[Optimalisasi Blended Learning Model Flipped Classroom pada Perkuliahan Time Series di Prodi Matematika](#)

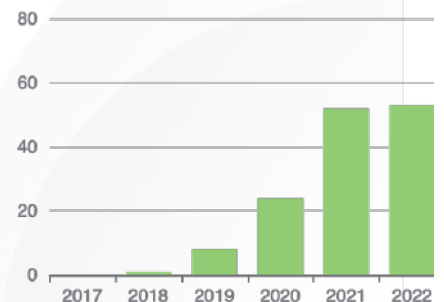
Prodi Pendidikan matematika Universitas Islam Majapahit (UNIM), Mojokerto, Indonesia
& [MAJAMATH: Jurnal Matematika dan Pendidikan Matematika Vol. 5 No. 1 \(2022\): Vol. 5 No. 1 Maret 2022 25 - 34](#)

 2022 [DOI: 10.36815/majamath.v5i1.1756](#) [Accred : Sinta 4](#)

[Estimation of Logistic Regression Model on The Risk of Acute Respiratory Infection Toddlers at Cikampak Torgamba Public Health Center Using The Maximum Likelihood Method](#)

Prodi Pendidikan matematika Universitas Islam Majapahit (UNIM), Mojokerto, Indonesia
& [MAJAMATH: Jurnal Matematika dan Pendidikan Matematika Vol. 5 No. 1 \(2022\): Vol. 5 No. 1 Maret 2022 35 - 42](#)

Citation Per Year By Google Scholar



Journal By Google Scholar

	All	Since 2020
Citation	302	288
h-index	9	9
i10-index	9	9

Home / Editorial Team

Editorial Team

Ketua Penyunting
[Suesthi Rahayuningsih](#) (Universitas Islam majapahit, Mojokerto, Indonesia)

Penyunting Pelaksana
[Feriyanto](#) (Universitas Islam Majapahit, Mojokerto, Indonesia)
[Ulii Nurul Imanah](#) (Universitas Islam Majapahit, Mojokerto, Indonesia)
[Kamirsyah Wahyu](#) (Universitas Islam Negeri Mataram, Mataram, Indonesia)
[Wilda Syam Tonra](#) (Universitas Khairun Ternate, Maluku Utara, Indonesia)
[Endang Suprpti](#) (Universitas Muhammadiyah Surabaya, Surabaya, Indonesia)
[Moch. Lutfianto](#) (STKIP Al Hikmah, Surabaya, Indonesia)

Reviewer
[Agustan Syamsuddin](#) ([Scopus ID: 57194533129](#)) Universitas Muhammadiyah Makassar, Makassar, Indonesia
[Iis Holisin](#) ([Scopus ID: 57204905955](#)) Universitas Muhammadiyah Surabaya, Surabaya, Indonesia
[Imam Rofiki](#) ([Scopus ID: 57200654458](#)) Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia
[Andi Mulawakkan Firdaus](#) ([Scopus ID: 36674946300](#)) Universitas Muhammadiyah Makassar, Makassar, Indonesia
[Hasan Basri](#) (Universitas Madura, Pamekasan, Indonesia)
[Deka Anjariyah](#) (Universitas Islam Majapahit, Mojokerto, Indonesia)
[Hendra Cipta](#) (Universitas Islam Negeri Sumatera Utara Medan, Indonesia)
[Andi Harpeni Dewantara](#) (IAIN Bone, Sulawesi Selatan, Indonesia)
[Sunyoto Hadi Prajitno](#) (Universitas PGRI Adi Buana Surabaya)

KEBIJAKAN JURNAL
Fokus dan Ruang Lingkup
Etika Publikasi
Akses Terbuka
Frekuensi Terbitan
Proses Penyuntingan
Biaya Naskah
Plagiarism

NASKAH
Panduan Penulisan
Penyerahan Online
Hak Cipta
Privasi
Template
Reviewer



[View My Stats](#)



SIMULASI *CLUSTERING* PROVINSI DI INDONESIA DALAM PENYEBARAN COVID-19 BERDASARKAN INDIKATOR KESEHATAN MASYARAKAT MENGGUNAKAN ALGORITMA *GAUSSIAN MIXTURE MODEL*

Joko Riyono¹, Sofia Debi Puspa², Christina Eni Pujiastuti³
^{1,2,3}Universitas Trisakti, Jl.Kyai Tapa No.1 Grogol, Jakarta
¹jokoriyono@trisakti.ac.id

Abstrak

Gaussian Mixture Model adalah suatu metode yang mengkonstruksikan *Clustering* suatu dataset menjadi beberapa kelompok data yang memiliki distribusi Gaussian atau Normal. Pada penelitian ini akan dibahas gagasan untuk menentukan *Clustering* penyebaran Covid-19 pada 34 provinsi di Indonesia menggunakan *Gaussian Mixture Model* berdasarkan nilai indikator kesehatan masyarakat. Mengingat masih berlangsungnya pandemi Covid-19 di beberapa provinsi di Indonesia hingga saat ini, penelitian ini dipilih dengan tujuan sebagai masukan kepada pemerintah selaku pembuat kebijakan untuk bahan acuan penanganan pandemi Covid-19 sehingga program-program pencegahan penyebaran Covid-19 di tiap provinsi dapat tertangani secara lebih optimal. Hasil analisis data diperoleh 6 cluster optimal yaitu cluster yang berpotensi sangat tinggi berisi 1 provinsi, tinggi 4 provinsi, sedang 13 provinsi, cukup rendah 4 provinsi, rendah 2 provinsi, sangat rendah 10 provinsi dalam penyebaran Covid-19.

Kata Kunci: *Gaussian Mixture Model*; *Clustering*; Indikator Kesehatan Masyarakat

Abstract

An Gaussian Mixture Model is a method that constructs clustering a dataset into several data groups that have a Gaussian or Normal distribution. This study will discuss the idea of determining the clustering of the spread of Covid-19 in 34 provinces in Indonesia using the Gaussian Mixture Model based on the value of public health indicators. Considering the ongoing Covid-19 pandemic in several provinces in Indonesia to date, this research was chosen with the aim of being an input to the government as a policy maker for reference material for handling the Covid-19 pandemic so that programs to prevent the spread of Covid-19 in each province can be handled more optimally. The results of data analysis obtained 6 optimal clusters, namely a very high potential cluster containing 1 province, high 4 provinces, medium 13 provinces, quite low 4 provinces, low 2 provinces, very low 10 provinces in the spread of Covid-19.

Keywords: Gaussian Mixture Model; Clustering; Public Health Indicators

Pendahuluan

Virus SARS-CoV-2 merupakan penyebab penyakit Covid 19. Virus ini merupakan salah satu jenis *coronavirus* yang masuk ke dalam satu kelompok virus penyebab penyakit SARS, yang dulu pernah terjadi pada tahun 2003 dan menewaskan sekitar 774 jiwa. Menurut penelitian yang

dilakukan (Olivia,2020) penyakit Covid-19 sendiri setidaknya 10 kali lebih menular jika dibandingkan dari penyakit SARS. Penderita penyakit Covid-19 dapat mengalami demam, batuk kering, dan juga kesulitan bernapas. Keluhan yang diderita bahkan dapat berkembang menjadi *pneumonia* berat, kegagalan multiorgan bahkan kematian. Dilansir dari <https://covid19.go.id/>, sejak 2 Maret 2020 hingga 8 September 2021 telah terjadi 4.147.365 total kasus penyakit Covid-19 di Indonesia dengan menewaskan 137.782 korban jiwa. Pemerintah pusat maupun daerah pun hingga saat ini berupaya menanggulangi atau mengurangi penyebaran penyakit Covid-19 dengan mencanangkan berbagai program. Menurut (Anuraga, 2015), strategi perencanaan program seperti kewajiban memakai masker, mempraktikkan aktifitas jaga jarak, memberikan program vaksinasi, dan lain-lain bisa saja berbeda untuk tiap daerah bergantung pada pola demografi epidemi yang ditemui dan terjadi.

Mengingat masih berlangsungnya pandemi Covid-19 di beberapa provinsi dan upaya pemerintah dalam memberantas Covid-19 juga masih berlangsung, membuat keinginan untuk memilih topik identifikasi *cluster* penyebaran Covid-19 di setiap provinsi di Indonesia berdasarkan indikator kesehatan masyarakat menggunakan algoritma *Gaussian Mixture Model* berbantuan *software RStudio*. Diharapkan hasil yang diperoleh dalam penelitian ini dapat memberi masukan kepada pemerintah selaku pembuat kebijakan untuk bahan acuan penanganan pandemi Covid-19 sehingga program-program pencegahan penyebaran Covid-19 di tiap provinsi dapat tertangani secara lebih optimal.

Hasil penelitian (Diah NS & Irma Y, 2020) memberikan gambaran umum mengenai tingkat keparahan serta resiko penyebaran covid-19 pada 34 provinsi di Indonesia melalui *cluster* yang terbentuk menggunakan metode *K- Means* berdasarkan variabel *Case Fatality Rate*, Total Kasus dan kepadatan penduduk. Hasil yang didapat 3 *cluster* dengan provinsi Jakarta dan Jawa Timur memiliki tingkat keparahan yang tinggi. Adanya Indikator kesehatan masyarakat yang telah ditetapkan pemerintah sebagai patokan untuk penilaian risiko penularan Covid-19 di suatu wilayah mendorong penelitian ini untuk mengambil nya sebagai variabel dalam menentukan clustering 34 provinsi di Indonesia menggunakan *Gaussian Mixture Model*.

Algoritma *Gaussian Mixture Model* merupakan metode analisis *cluster non-hierarchical* yang bekerja untuk memodelkan beberapa data dalam distribusi Gaussian dengan parameter mean dan varians tertentu (Putra, 2017). Dalam melakukan *clustering*, Algoritma *Gaussian Mixture Model* adalah salah satu jenis *soft clustering* dimana satu *data point* bisa berada pada dua atau lebih *cluster*. Menurut (Amaluddin, 2015) Algoritma *Gaussian Mixture Model* sendiri merupakan model statistik yang sangat populer dan umum digunakan. Selain itu, algoritma ini

relatif efisien untuk pengolahan objek dalam jumlah besar. Berdasarkan hasil penelitian (Amaluddin, 2015) dan alasan bahwa tidak selalu *data point* tepat berada pada satu *cluster* tetapi bisa jadi berada di dua atau lebih *cluster* serta memperhitungkan besarnya probabilitas data point berada dalam suatu *cluster*, maka algoritma *Gaussian Mixture Model* akan digunakan dalam penelitian ini.

Dalam melakukan *clustering* menggunakan *Gaussian Mixture Model*, penentuan jumlah *cluster* yang paling sesuai untuk dataset yang dianalisa merupakan masalah yang cukup kompleks. Untuk dapat tercapainya proses ini keseimbangan antara kesempurnaan suatu model dengan tepat tidaknya suatu model terhadap data yang dianalisa harus benar-benar terjaga. Dengan kata lain, model yang dipilih untuk suatu dataset haruslah cukup sempurna agar dapat mencakup semua informasi yang terkandung di dalam dataset tersebut, tetapi tidak harus selalu sempurna agar tidak mengalami *over-fit*. (Luca Scrucca, 2016) menyatakan bahwa pada *Gaussian Mixture Model*, *cluster* yang terbentuk berbentuk *ellipsoidal* dengan pusat di μ_k . Matrik *Covarians* Σ_k menentukan karakteristik geometri yaitu *shape*, *volume* dan orientasi. Karakteristik geometrik tersebut dapat dibuat beragam antar *cluster* atau dibuat sama. Interpretasi geometrik dari berbagai parameterisasi matrik *covarians* Σ_k sebaran peluang *Gauss* pada *package Mclust* di *RStudio* dapat terlihat pada Tabel 1. Adapun jumlah *cluster* optimal dapat ditentukan berdasarkan pemilihan model terbaik. (Fraley, 2012) Pendekatan yang umum dipakai dalam menentukan pemilihan model terbaik adalah menggunakan *Bayes Information Criterion* (BIC). Nilai BIC diperoleh menggunakan rumus :

$$\text{BIC} \approx 2 \log p(x/M) + \text{konstanta} \approx 2l_M(x, \hat{\theta}) - m_M \log(n) \quad (1)$$

Dimana :

$p(x/M)$ = likelihood dari data untuk model M

$L_M(x, \hat{\theta})$ = loglikelihood mixture maksimum untuk model M

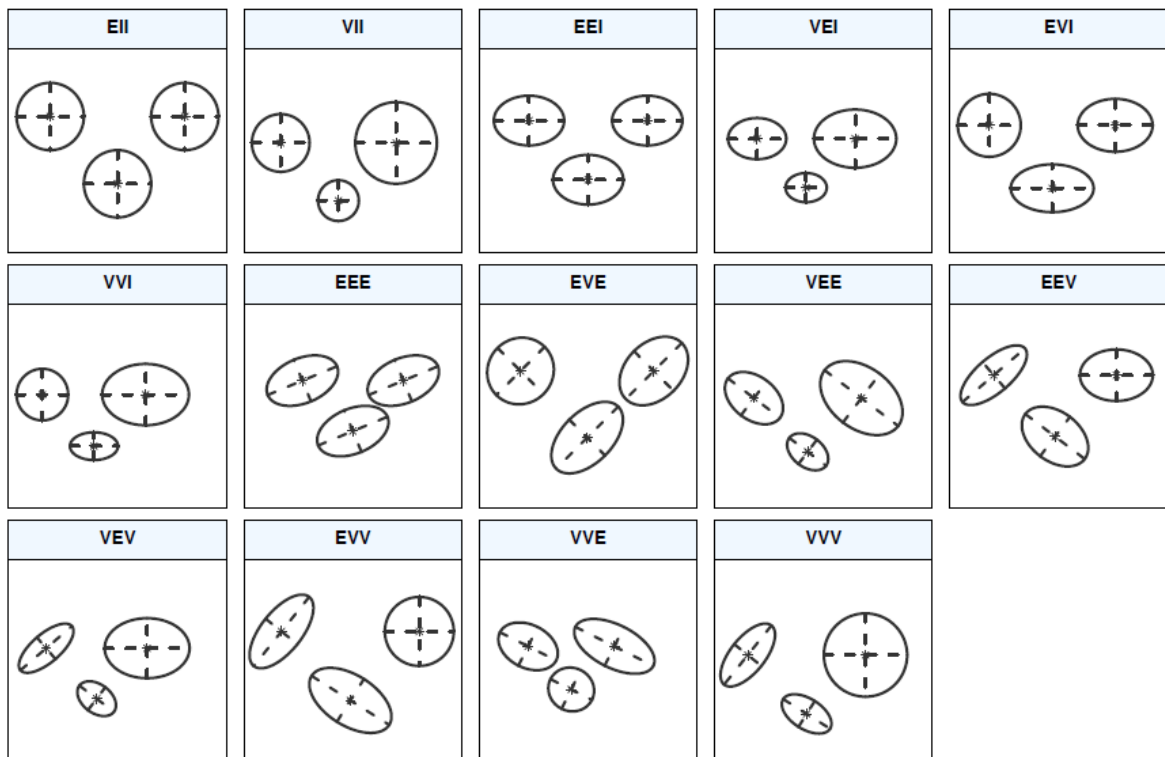
m_M = jumlah parameter bebas yang diduga dalam model

Model terbaik dipilih berdasarkan nilai BIC terbesar. Nilai BIC dapat dihitung menggunakan *RStudio* dengan *package Mclust*.

Tabel 1. Parameterisasi matriks covarians dalam grup Σ_k untuk data multidimensi tersedia dalam paket mclust, dan karakteristik geometris yang sesuai

Model	Σ_k	Distribution	Volume	Shape	Orientation
EII	λI	Spherical	Equal	Equal	—
VII	$\lambda_k I$	Spherical	Variable	Equal	—
EEI	λA	Diagonal	Equal	Equal	Coordinate axes
VEI	$\lambda_k A$	Diagonal	Variable	Equal	Coordinate axes
EVI	λA_k	Diagonal	Equal	Variable	Coordinate axes
VVI	$\lambda_k A_k$	Diagonal	Variable	Variable	Coordinate axes
EEE	$\lambda D A D^T$	Ellipsoidal	Equal	Equal	Equal
EVE	$\lambda D A_k D^T$	Ellipsoidal	Equal	Variable	Equal
VEE	$\lambda_k D A D^T$	Ellipsoidal	Variable	Equal	Equal
VVE	$\lambda_k D A_k D^T$	Ellipsoidal	Variable	Variable	Equal
EEV	$\lambda D_k A D_k^T$	Ellipsoidal	Equal	Equal	Variable
VEV	$\lambda_k D_k A D_k^T$	Ellipsoidal	Variable	Equal	Variable
EVV	$\lambda D_k A_k D_k^T$	Ellipsoidal	Equal	Variable	Variable
VVV	$\lambda_k D_k A_k D_k^T$	Ellipsoidal	Variable	Variable	Variable

Sumber: mclust 5 Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models by Luca Scrucca, Michael Fop, T. Brendan Murphy and Adrian E. Raftery



Gambar 1. Elips isodensitas untuk 14 model Gaussian yang diperoleh dengan dekomposisi eigen dalam kasus tiga kelompok dalam dua dimensi.

Sumber: mclust 5 Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models by Luca Scrucca, Michael Fop, T. Brendan Murphy and Adrian E. Raftery

Gaussian Mixture Model adalah suatu metode yang memodelkan atau mengelompokkan (*clustering*) data suatu dataset menjadi beberapa kelompok data yang memiliki distribusi probabilitas Gaussian atau Normal. Metode ini mengasumsikan bahwa keseluruhan individu adalah campuran dari sebaran peluang Gaussian, mewakili distribusi Gaussian dimana masing

masing sebaran secara khas mempunyai parameter distribusi. Salah satu metode yang digunakan untuk menduga parameter adalah melalui *algoritma Expectation Maximization* (EM). Algoritma EM merupakan algoritma iteratif populer yang dapat digunakan untuk menemukan penduga parameter dengan memaksimalkan *loglikelihood*.

EM dimulai dengan inisialisasi nilai awal dengan parameter *mixture model*, kemudian secara iteratif memperbaharui bagian parameternya. Inisialisasi nilai awal diperoleh menggunakan *algoritma hierarchical Clustering* sedang banyaknya cluster ditentukan menggunakan *Bayes Information Criteria (BIC)*, model terbaik dengan nilai BIC paling maksimal (Frale, 2012).

Algoritma *Expectation Maximization* adalah salah satu alternatif algoritma yang banyak digunakan dalam melakukan pemodelan *mixture*. Adapun langkah-langkah *Gaussian Mixture* dengan EM Algoritma sebagai berikut :

- 1) Inisialisasi dengan m Gaussian secara acak

$$(\mu_a, \sigma_a^2), (\mu_b, \sigma_b^2), \dots, (\mu_m, \sigma_m^2)$$

- 2) Ulangi hingga konvergen

- a) E-Step apakah x_i terlihat masuk di cluster a, b, c, ..., m

$$a_i = a = P(a|x_i) = \frac{P(x_i|a) \cdot P(a)}{P(x_i)}, i = 1, 2, \dots, n$$

$$b_i = b = P(b|x_i) = \frac{P(x_i|b) \cdot P(b)}{P(x_i)}, i = 1, 2, \dots, n$$

.

.

.

$$m_i = m = P(m|x_i) = \frac{P(x_i|m) \cdot P(m)}{P(x_i)}, i = 1, 2, \dots, n \quad (2)$$

- b) M-Step perbaiki $(\mu_a, \sigma_a^2), (\mu_b, \sigma_b^2), \dots, (\mu_m, \sigma_m^2)$ sampai konvergen

$$\mu_a = \frac{a_1 x_1 + a_2 x_2 + \dots + a_n x_n}{a_1 + a_2 + \dots + a_n}$$

$$\sigma_a^2 = \frac{a_1 (x_1 - \mu_a)^2 + a_2 (x_2 - \mu_a)^2 + \dots + a_n (x_n - \mu_a)^2}{a_1 + a_2 + \dots + a_n}$$

.

.

.

$$\mu_m = \frac{m_1 x_1 + m_2 x_2 + \dots + m_n x_n}{m_1 + m_2 + \dots + m_n}$$

$$\sigma_m^2 = \frac{m_1(x_1 - \mu_a)^2 + m_2(x_2 - \mu_a)^2 + \dots + m_n(x_n - \mu_a)^2}{m_1 + m_2 + \dots + m_n} \quad (3)$$

$$P(a)_{baru} = \frac{a_1 + a_2 + \dots + a_n}{n}$$

$$P(b)_{baru} = \frac{b_1 + b_2 + \dots + b_n}{n}$$

.

.

.

$$P(m)_{baru} = \frac{m_1 + m_2 + \dots + m_n}{n} \quad (4)$$

Setelah didapat nilai $(\mu_a, \sigma_a^2), (\mu_b, \sigma_b^2), \dots, (\mu_m, \sigma_m^2)$ konvergen, hitung probabilitas cluster setiap data dari dataset berada, nilai probabilitas digunakan untuk menentukan cluster data dari dataset dengan cara melihat nilai yang terbesar.

Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah penelitian studi literatur dimana data siap pakai tentang jumlah penduduk dan data kasus covid-19 baik berupa kasus positif, sembuh dan meninggal akibat covid-19 tiap provinsi diambil dari *database* situs <https://kawalcovid19.id/> dan *database* situs resmi Kementrian Dalam Negeri Indonesia. Data yang digunakan adalah data penyebaran dari tanggal 15 Maret 2020 hingga 29 Oktober 2021. Berdasarkan data ini akan dibuat *clustering* penyebaran Covid-19 di setiap provinsi di Indonesia berdasarkan indikator kesehatan masyarakat menggunakan algoritma *Gaussian Mixture Model* berbantuan *software RStudio*. Adapun tahapan penelitian ini dilakukan sebagai berikut:

Data yang dibutuhkan dalam penelitian ini adalah data kasus Covid-19 yang diperoleh dari *database* situs <https://kawalcovid19.id/>, data jumlah penduduk yang diperoleh dari *database* situs resmi Kementrian Dalam Negeri Indonesia. Data kasus Covid-19 yang digunakan berisikan informasi berupa nilai kasus positif, sembuh, dan meninggal akibat Covid-19 per-provinsi dari 15 Maret 2020 hingga 29 Oktober 2021.

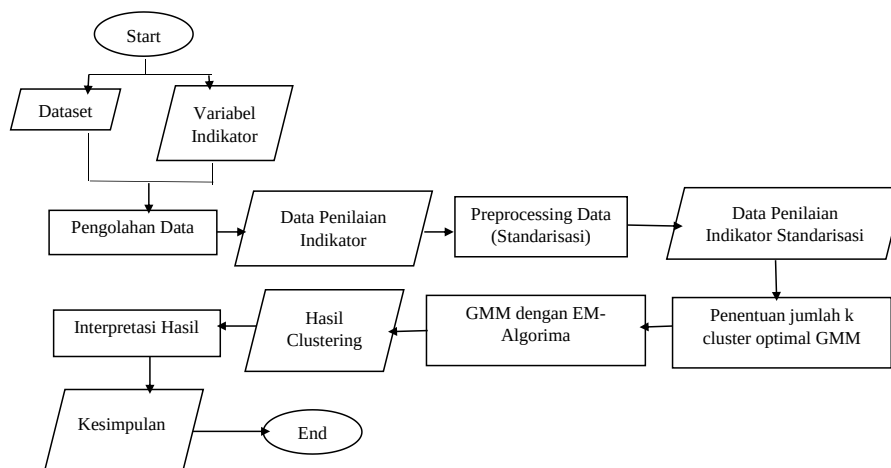
Analisis kebutuhan dilakukan dengan menentukan kebutuhan apa saja yang diperlukan dalam implementasi algoritma *Gaussian Mixture Model* untuk *clustering* daerah penyebaran Covid-19. Pada penelitian ini akan ditentukan beberapa indikator yang akan digunakan untuk menilai situasi Covid-19 di masyarakat dengan mengambil 5 indikator kesehatan masyarakat yang telah ditetapkan sebelumnya oleh pemerintah Indonesia, untuk penanganan pandemi Covid-19.

Standarisasi merupakan salah satu proses processing dengan menyamakan satuan dari masing-masing atribut. Dalam kasus dimana data memiliki rentang nilai yang jauh berbeda dapat

menyebabkan perhitungan menjadi tidak efektif (Pandie,2018). Jadi standarisasi dilakukan agar data memiliki rentang nilai yang sama sehingga diharapkan data dengan nilai yang besar atau kecil tidak akan mempengaruhi hasil analisis akhir. Salah satu cara menstandarisasi data adalah dengan *Z-Score*, pada dasarnya menstandarisasi menggunakan *Z-Score* adalah mengubah nilai data asli menjadi bentuk Z (data berdistribusi Normal) dengan rumus :

$$z = \frac{x_i - \bar{x}}{s} \text{ dimana } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ dan } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

Perancangan penelitian merupakan rancangan langkah kerja secara keseluruhan, yang dilakukan agar dapat mempermudah implementasi. Pada penelitian ini akan dilakukan berdasarkan langkah kerja pada Gambar 2 sebagai berikut:



Gambar 2. Perancangan Sistem

Pada Gambar 2 terlihat diagram alir penelitian ini menggunakan algoritma Gaussian Mixture Model. Diawali dengan melakukan pengumpulan data, data yang digunakan pada penelitian ini adalah data kasus Covid-19 harian setiap provinsi di wilayah Indonesia dan penentuan variabel-variabel indikator kesehatan masyarakat yaitu rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif yang terjadi di tiap provinsi, rata-rata persentase penurunan angka kematian, rata-rata kenaikan jumlah sembuh dari kasus positif, rata-rata persentase penurunan jumlah meninggal dari kasus positif selama 2 minggu terakhir dari puncak kasus meninggal akibat covid-19 yang terjadi di tiap provinsi, rata-rata persentase penurunan laju insidensi kasus positif dimana dari kelima indikator tersebut semakin besar nilainya maka semakin kecil potensi penyebaran Covid-19 pada daerah tersebut. Dari sini akan dilakukan perhitungan untuk mendapatkan data nilai variabel indikator kesehatan masyarakat. Selanjutnya dilakukan preprocessing data dengan standarisasi agar data memiliki

rentang nilai yang sama sehingga diperoleh data nilai variabel indikator kesehatan masyarakat yang telah terstandarisasi. Proses selanjutnya melakukan perhitungan BIC untuk menentukan jumlah *cluster* optimal diikuti dengan analisis *Clustering* dengan *Gaussian Mixture Model* untuk mendapatkan hasil *Clustering* provinsi di Indonesia berdasarkan indikator kesehatan masyarakat. Pada langkah terakhir dilakukan penarikan kesimpulan dengan menentukan *cluster* dan provinsi yang masuk di dalam *cluster* berdasarkan nilai probabilitasnya.

Hasil dan Pembahasan

Berdasarkan data mentah Covid-19 pada 34 provinsi di Indonesia per hari mulai 15 Maret 2020 sampai 29 Oktober 2021 yaitu kasus positif, sembuh, dan meninggal akibat Covid-19 serta data jumlah penduduk akan dihitung 5 nilai variabel indikator kesehatan masyarakat yaitu: Rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif terjadi di setiap provinsi (X_1)

$$\% \text{ kasus positif} = \frac{(\text{Jumlah kasus Covid-19 pada } t - \text{Jumlah kasus Covid-19 pada } t+1)}{\text{Jumlah kasus Covid-19 pada } t} \times 100\% \quad (6)$$

Rata-rata persentase penurunan jumlah kematian dari kasus positif selama 2 minggu terakhir dari puncak kasus meninggal akibat Covid-19 terjadi di setiap provinsi (X_2)

$$\% \text{ kasus kematian} = \frac{(\text{Jumlah kasus Covid-19 pada } t - \text{Jumlah kasus Covid-19 pada } t+1)}{\text{Jumlah kasus Covid-19 pada } t} \times 100\% \quad (7)$$

Rata-rata persentase kenaikan jumlah sembuh dari kasus positif (X_3)

$$\% = \frac{(\text{Jumlah kasus sembuh Covid-19 pada } t+1 - \text{Jumlah kasus sembuh Covid-19 pada } t)}{\text{Jumlah kasus sembuh Covid-19 pada } t} \times 100\% \quad (8)$$

Rata-rata persentase penurunan laju insidensi kasus positif (X_4)

$$LI = \frac{\text{jumlah kasus positif Covid 19 pada suatu periode waktu}}{\text{jumlah waktu (person-time at risk) selama periode tersebut}} \quad (9)$$

Rata-rata persentase penurunan angka kematian (X_5)

$$CFR = \frac{\text{jumlah kasus meninggal akibat Covid 19 kumulatif}}{\text{jumlah kasus positif Covid 19 kumulatif}} \times k \quad (10)$$

Untuk $k=100.000$ artinya nilai CFR yang didapat merupakan angka kematian per 100.000 kasus positif.

Ke lima nilai variabel indikator kesehatan masyarakat yaitu : Rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif terjadi di setiap provinsi (X_1), Rata-rata persentase penurunan jumlah kematian dari kasus positif selama 2 minggu

terakhir dari puncak kasus meninggal akibat Covid-19 terjadi di setiap provinsi (X_2), Rata-rata persentase kenaikan jumlah sembuh dari kasus positif (X_3), Rata-rata persentase penurunan laju insidensi kasus positif (X_4), dan Rata-rata persentase penurunan angka kematian (X_5) di atas untuk tiap provinsi disajikan di dalam tabel 2 berikut ini:

Tabel 2 Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi

i	Provinsi	Variabel Penelitian				
		X_1	X_2	X_3	X_4	X_5
1	Aceh	-0.017395058	-0.12567459	0.47180616	-0.350422889	0.001146201
2	Bali	0.030149687	0.033427009	0.24807565	-0.173669117	0.0023349
3	Banten	0.004932763	-0.107142857	0.801298204	-0.457009101	-0.001158132
4	Bangka	-0.095474577	-0.010527865	0.146223155	-0.122875531	0.002839782
	Belitung					
5	Bengkulu	-0.076193964	-0.098095238	0.071668478	-0.589690286	0.003816346
6	DIY	0.000270674	-0.027748812	0.189858158	-0.209640353	0.001097903
	Jogjakarta					
7	Jakarta	0.08402681	-0.319930168	0.431768005	-0.072633559	0.004271964
8	Jambi	-1.336752996	-0.147340398	0.207678834	-0.256789831	-0.003196558
9	Jabar	-0.017359225	-0.675228245	0.436483966	-0.234587756	0.004815199
10	Jateng	-0.004614816	0.019290123	0.869659053	-0.234258239	0.002299963
11	Jatim	-0.025512409	0.045263981	0.113976158	-0.18890274	-0.001524052
12	Kalbar	-0.241148184	-0.257538592	0.064887385	-0.062881213	0.002506188
13	Kaltim	-0.007455939	0.011884017	0.241373218	-0.047787712	-0.000963926
14	Kalteng	0.037252511	-0.6555553	0.427742266	-0.138285453	-0.000207034
15	Kalsel	-0.029724393	0.05853192	0.291360409	-0.365975128	-0.002311997
16	Kalut	-0.022592743	-0.037221112	0.993920866	-0.281131215	0.000196384
17	Kep. Riau	0.039153866	0.081211179	1.589752179	-0.192996638	-0.000821704
18	NTB	-0.023672621	0.214285714	1.178139238	-0.580371172	0.000893519
19	Sumsel	-0.145434791	-0.168967516	0.404936954	-0.186154992	0.001922869
20	Sumbar	0.02659528	-0.060716065	0.287237516	-0.256454451	0.001743815
21	Sulut	-0.184530382	-0.771910431	1.337859852	-0.739156669	0.00150216
22	Sumut	-0.018213249	-0.543677912	0.054271858	-0.201746631	0.002265546
23	Sulteng	-0.040592164	-0.005470522	1.292578376	-0.811774395	0.000178654
24	Sulsel	-0.002795466	-0.033684119	0.306069802	-0.303767012	0.000831231
25	Sulteng	-0.051019175	0.024042525	0.3968188	-0.365563194	0.002826596
26	Lampung	0.056055904	0.053954816	0.248494695	-0.157278112	-0.002400012
27	Riau	0.018812508	0.024352882	0.135039888	-0.133431093	0.001912152
28	Maluku Utara	-0.075544727	0.058967223	0.586915878	-0.526094572	0.000173624
29	Maluku	-0.029569199	-0.222222222	0.536073196	-0.611446853	0.001788984
30	Papua Barat	-0.328061018	0.071402407	0.902740251	-0.667480196	0.003314603
31	Papua	-0.163812042	-0.749285714	1.756314314	-0.576878443	-0.000121436
32	Sulbar	-0.117340052	0.045562771	0.841715538	-0.794864684	0.001404975
33	NTT	0.068506452	-0.189044414	1.525624921	-0.742004078	-6.53253E-05
34	Gorontalo	0.021139723	-0.147363946	0.592703378	-0.646223074	0.000409086

Proses selanjutnya Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi di Tabel 2 distandarisasi. Standarisasi merupakan salah satu proses preprocessing, hasil standarisasi data Tabel 2 Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi dengan RStudio disajikan seperti berikut ini:

```
##          x1          x2          x3          x4          x5
## [1,] 0.255279028 0.03902542 -0.240989472 0.04599271 0.08077947
## [2,] 0.454011397 0.66132919 -0.706300956 0.80149500 0.70283213
## [3,] 0.348607128 0.11150971 0.444283321 -0.40959093 -1.12509072
## [4,] -0.071085698 0.48940581 -0.918132302 1.01860307 0.96703913
## [5,] 0.009505371 0.14689822 -1.073190043 -0.97671285 1.47808048
## [6,] 0.329120056 0.42204848 -0.827380853 0.64774241 0.05550472
## [7,] 0.679212481 -0.72077824 -0.324260247 1.23335337 1.71650793
## [8,] -5.259507661 -0.04571737 -0.790317669 0.44621042 -2.19180952
## [9,] 0.255428807 -2.11047724 -0.314452060 0.54110921 2.00078555
## [10,] 0.308699183 0.60603472 0.586459227 0.54251767 0.68454899
## [11,] 0.221349299 0.70762786 -0.985199141 0.73638161 -1.31657846
## [12,] -0.679987079 -0.47674228 -1.087293261 1.27503802 0.79246764
## [13,] 0.296823569 0.57706677 -0.720240577 1.33955248 -1.02346126
## [14,] 0.483700502 -2.03352925 -0.332632923 0.95273613 -0.62737599
## [15,] 0.203743616 0.75952355 -0.616277937 -0.02048254 -1.72891415
## [16,] 0.233553217 0.38499889 0.844897150 0.34216755 -0.41626505
## [17,] 0.491647978 0.84823032 2.084098509 0.71888298 -0.94903596
## [18,] 0.229039431 1.36873285 1.228031859 -0.93687997 -0.05145026
## [19,] -0.279914489 -0.13030884 -0.380063080 0.74812637 0.48721405
## [20,] 0.439154322 0.29310165 -0.624852672 0.44764394 0.39351396
## [21,] -0.443330224 -2.48863615 1.560216486 -1.61558009 0.26705449
## [22,] 0.251859069 -1.59593641 -1.109371282 0.68148270 0.66653879
## [23,] 0.158317401 0.50918691 1.466040725 -1.92597153 -0.42554314
## [24,] 0.316303888 0.39883335 -0.585685555 0.24541489 -0.08404607
## [25,] 0.114733519 0.62462308 -0.396947100 -0.01872180 0.96013914
## [26,] 0.562296840 0.74162085 -0.705429431 0.87155541 -1.77497269
## [27,] 0.406623102 0.62583700 -0.941391100 0.97348522 0.48160541
## [28,] 0.012219119 0.76122618 -0.001585943 -0.70488438 -0.42817547
## [29,] 0.204392313 -0.33860721 -0.107327819 -1.06970739 0.41715125
## [30,] -1.043274193 0.80986467 0.655261026 -1.30921186 1.21551547
## [31,] -0.356729596 -2.40014272 2.430512030 -0.92195092 -0.58258165

## [32,] -0.162481238 0.70879653 0.528342710 -1.85369400 0.21619732
## [33,] 0.614338910 -0.20883683 1.950727564 -1.62775082 -0.55321885
## [34,] 0.416350631 -0.04580948 0.010450815 -1.21835211 -0.30495671
## attr(,"scaled:center")
##          x1          x2          x3          x4          x5
## -0.0784680298 -0.1356520432 0.5876784294 -0.3611831289 0.0009918373
## attr(,"scaled:scale")
##          x1          x2          x3          x4          x5
## 0.23924007 0.25566549 0.48081880 0.23395531 0.00191093
```

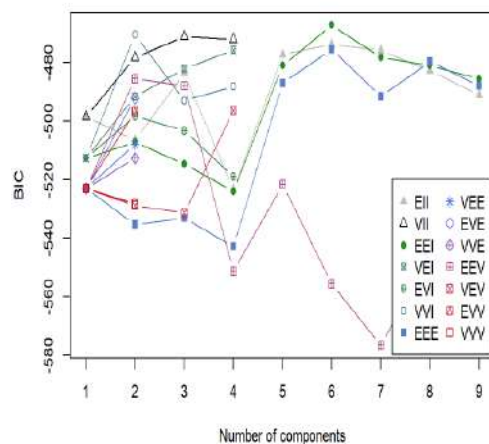
Gambar 3.Hasil standarisasi dengan RStudio Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi dengan RStudio

Interpretasi *output* diperoleh dari gambar 3 hasil standarisasi data nilai setiap variabel indikator kesehatan masyarakat disetiap provinsi dengan nilai *mean* dan *standard deviasinya* untuk setiap variabel seperti terangkum berikut : (1) Rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif terjadi di setiap provinsi (X_1) adalah -0,0784680298 dan 0,23924007, (2) Rata-rata persentase penurunan jumlah kematian dari kasus positif selama 2 minggu terakhir dari puncak kasus meninggal akibat Covid-19 terjadi di setiap provinsi (X_2) adalah -0,1356520432 dan 0,25566549, (3) Rata-rata persentase kenaikan jumlah sembuh dari kasus positif (X_3) adalah 0,5876784294 dan 0,48081880, (4) Rata-rata persentase penurunan laju insidensi kasus positif (X_4) adalah -0,3611831289 dan 0,23395531 dan (5) Rata-rata persentase penurunan angka kematian (X_5) adalah 0,0009918373 dan 0,00191093

Langkah berikutnya menentukan jumlah cluster dan model parameter optimal didasarkan nilai BIC, menggunakan Rstudio diperoleh:


```
## Bayesian Information Criterion (BIC):
##      EII      VII      EEI      VEI      EVI      VVI      EEE
## 1 -498.5223 -498.5223 -512.6277 -512.6277 -512.6277 -512.6277 -522.9730
## 2 -506.4091 -478.3174 -507.2429 -491.7012 -498.4362 -470.5172 -535.3390
## 3 -483.7178 -471.2158 -514.6682 -482.1353 -503.4178 -493.0061 -533.0706
## 4 -523.5063 -472.2403 -523.8824 -475.8752 -519.0186 -488.2190 -542.8600
## 5 -477.3656      NA -481.0950      NA      NA      NA -487.1784
## 6 -473.8922      NA -467.3299      NA      NA      NA -475.5765
## 7 -476.0110      NA -478.3207      NA      NA      NA -491.6081
## 8 -482.7750      NA -481.0566      NA      NA      NA -479.7325
## 9 -491.1532      NA -485.6091      NA      NA      NA -487.9913
##      VEE      EVE      VVE      EEV      VEV      EVV      VVV
## 1 -522.9730 -522.9730 -522.9730 -522.9730 -522.9730 -522.973 -522.9730
## 2 -508.0239 -492.1949 -512.7047 -485.5947 -528.8886 -496.808 -528.3212
## 3      NA      NA      NA -488.0284 -531.4230      NA      NA
## 4      NA      NA      NA -551.2766 -496.3652      NA      NA
## 5      NA      NA      NA -521.6032      NA      NA      NA
## 6      NA      NA      NA -555.5112      NA      NA      NA
## 7      NA      NA      NA -576.4611      NA      NA      NA
## 8      NA      NA      NA -545.9692      NA      NA      NA
## 9      NA      NA      NA -525.1399      NA      NA      NA
##
## Top 3 models based on the BIC criterion:
##      EEI,6      VVI,2      VII,3
## -467.3299 -470.5172 -471.2158
```

Gambar 4. Hasil Bayes Information Criterion Data Penelitian Terstandarisasi dengan Rstudio



Gambar 5. Plot Hasil Bayes Information Criterion Data Penelitian Terstandarisasi dan Jumlah Cluster dengan Rstudio

Interpretasi hasil BIC dengan package Mclust Rstudio terlihat bahwa nilai terbesar BIC sebesar -467,3299 didapat untuk Parameterisasi matriks *covarians* dalam grup matriks *Covarians* (Σ_k) adalah EEI dengan jumlah *cluster* 6. Selanjutnya akan dilakukan membuat *Cluster Gaussian Mixture Model* dengan jumlah *cluster* sebanyak 6 dan model parameter EEI menggunakan Rstudio diperoleh hasil sebagai berikut:

```
## -----
## Gaussian finite mixture model fitted by EM algorithm
## -----
## Mclust EEI (diagonal, equal volume and shape) model with 6 components:
##
## log-likelihood n df      BIC      ICL
##      -163.1377 34 40 -467.3299 -470.3897
##
## Clustering table:
##  1  2  3  4  5  6
## 13 10  4  1  4  2
##
## Mixing probabilities:
##      1      2      3      4      5      6
## 0.38003593 0.29880174 0.11356940 0.02941176 0.11935763 0.05882353
##
## Means:
##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## x1 0.1397490 0.1359029 0.4071261 -5.25950766 0.3206147 -0.4000299
## x2 0.2437372 0.5261606 -1.6470652 -0.04571737 0.6898752 -2.4443894
## x3 -0.5836773 0.9259166 -0.5277380 -0.79031767 -0.7460100 1.9953642
## x4 0.3393889 -0.8278377 0.8383199 0.44621042 0.7094755 -1.2687655
## x5 0.5495671 -0.2261366 0.9104198 -2.19180952 -1.4321319 -0.1577636
##
## Variances:
## [,,1]
##      x1      x2      x3      x4      x5
## x1 0.1035401 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1488556 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2343173 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4815988 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3588329
## [,,2]
##      x1      x2      x3      x4      x5
## x1 0.1035401 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1488556 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2343173 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4815988 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3588329
## [,,3]
##      x1      x2      x3      x4      x5
## x1 0.1035401 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1488556 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2343173 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4815988 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3588329
## [,,4]
##      x1      x2      x3      x4      x5
## x1 0.1035401 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1488556 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2343173 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4815988 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3588329
## [,,5]
##      x1      x2      x3      x4      x5
## x1 0.1035401 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1488556 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2343173 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4815988 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3588329
## [,,6]
##      x1      x2      x3      x4      x5
## x1 0.1035401 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1488556 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2343173 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4815988 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3588329
```

Gambar 6. Hasil Clustering dengan GMM

Berdasarkan output di atas diperoleh hasil bahwa jumlah observasi sebanyak 34 ($n = 34$) diperoleh *clustering* GMM dengan jumlah *cluster* sebanyak 6 dan model parameter optimal adalah EEI, dimana untuk *cluster* 1 berisi 13 anggota, *cluster* 2 berisi 10 anggota, *cluster* 3 berisi 4 anggota, *cluster* 4 berisi 1 anggota, *cluster* 5 berisi 4 anggota, *cluster* 6 berisi 2 anggota. Berdasarkan output di atas diperoleh pula parameter sebaran Gaussian untuk masing-masing *cluster* sebagai berikut:

Untuk cluster 1,

$$C_1 \sim N \left(\mu_1 = \begin{pmatrix} 0.1397490 \\ 0.2437372 \\ -0.5836773 \\ 0.3393889 \\ 0.5495671 \end{pmatrix}, \sigma_1^2 = \begin{pmatrix} 0.1035401 \\ 0.1488556 \\ 0.2343173 \\ 0.4815988 \\ 0.3588329 \end{pmatrix} \right)$$

Dan seterusnya, pada cluster 6: $C_6 \sim N \left(\mu_6 = \begin{pmatrix} -0.4000299 \\ -2.4443894 \\ 1.851874201 \\ -1.535068823 \\ -0.323572286 \end{pmatrix}, \sigma_6^2 = \begin{pmatrix} 0.13061732 \\ 0.8769211 \\ 1.9953642 \\ -1.2687655 \\ -0.1577636 \end{pmatrix} \right)$

dapat diperoleh juga nilai probabilitas *cluster* setiap data dari dataset untuk menentukan cluster data dari data point dengan cara melihat nilai probabilitas terbesar, dengan hasil sebagai berikut:

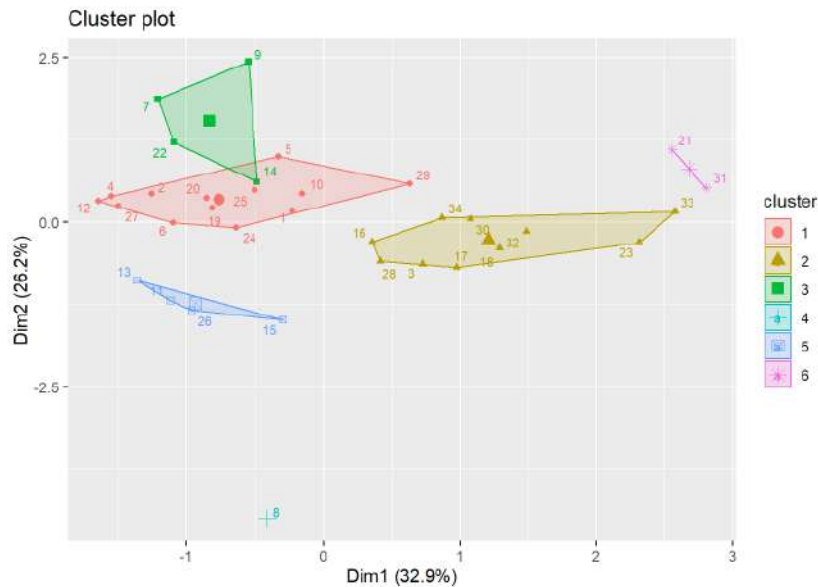
##	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
## [1,]	9.804585e-01	1.694791e-02	7.442415e-06	9.484159e-69	2.586143e-03	1.671277e-16
## [2,]	9.979986e-01	1.148552e-04	1.712992e-08	1.379142e-75	1.886542e-03	1.119737e-25
## [3,]	1.446738e-02	9.661449e-01	1.223603e-08	4.722505e-69	1.938772e-02	3.817387e-14
## [4,]	9.998498e-01	7.215096e-06	6.069754e-08	2.516135e-64	1.428880e-04	1.075828e-24
## [5,]	9.999384e-01	5.954786e-05	1.180557e-06	1.090328e-67	8.431393e-07	1.823940e-20
## [6,]	9.746863e-01	1.949395e-04	1.092151e-07	2.200377e-70	2.511867e-02	2.503171e-23
## [7,]	1.417624e-01	5.407514e-07	8.582371e-01	7.708968e-84	2.498250e-08	7.035004e-16
## [8,]	1.785161e-65	4.794725e-67	3.461269e-77	1.000000e+00	1.352383e-66	2.372094e-69
## [9,]	1.656559e-08	1.450708e-14	1.000000e+00	5.160753e-81	2.640470e-18	3.765981e-10
## [10,]	5.480674e-01	4.516788e-01	1.338973e-08	5.769333e-73	2.538043e-04	1.298365e-18
## [11,]	8.371290e-03	1.645351e-05	5.035902e-12	1.268912e-65	9.916123e-01	4.303786e-27
## [12,]	9.972012e-01	3.132144e-07	2.788717e-03	4.725632e-49	9.805278e-06	8.897969e-17
## [13,]	4.014928e-02	4.114015e-05	4.110614e-10	1.470673e-67	9.598096e-01	2.001943e-25
## [14,]	2.238252e-07	2.183038e-11	9.999998e-01	2.440776e-76	1.917167e-10	1.434418e-08
## [15,]	1.748963e-03	6.270966e-04	1.719345e-13	8.639816e-65	9.976239e-01	6.366632e-25
## [16,]	1.930858e-02	9.794087e-01	1.840910e-09	5.460168e-69	1.282688e-03	1.078239e-15
## [17,]	2.286122e-06	9.999874e-01	1.386343e-15	8.095511e-78	1.033533e-05	5.178517e-18
## [18,]	2.571785e-05	9.999731e-01	3.861337e-18	4.505327e-74	1.231873e-06	6.858899e-23
## [19,]	9.994355e-01	3.915926e-04	4.903659e-05	1.041400e-57	1.239107e-04	5.425778e-16
## [20,]	9.971147e-01	4.481276e-04	9.146827e-07	5.626932e-74	2.436226e-03	1.373048e-21
## [21,]	3.329172e-17	2.039869e-14	1.191025e-09	1.008824e-68	3.893277e-25	1.000000e+00
## [22,]	4.588793e-05	8.213376e-12	9.999541e-01	3.552717e-73	1.046972e-10	1.384698e-13
## [23,]	1.121822e-06	9.999989e-01	2.008347e-15	4.791671e-72	1.144034e-08	1.860727e-14
## [24,]	9.662613e-01	3.153352e-03	8.550362e-08	9.600726e-70	3.058526e-02	2.916629e-21
## [25,]	9.965666e-01	3.336548e-03	6.775268e-09	6.358995e-69	9.683606e-05	3.581895e-22
## [26,]	3.634320e-04	8.826697e-06	2.662998e-13	2.632589e-73	9.996277e-01	5.823714e-28
## [27,]	9.928897e-01	2.510760e-05	2.138196e-08	1.206154e-73	7.085204e-03	1.777305e-26
## [28,]	1.416374e-01	8.414802e-01	1.752640e-11	3.631820e-63	1.688237e-02	8.834618e-20
## [29,]	8.717564e-01	1.278828e-01	3.197523e-04	1.329126e-68	4.102182e-05	1.662431e-11
## [30,]	1.731589e-02	9.826841e-01	2.466900e-13	5.509252e-46	1.696255e-08	4.416188e-16
## [31,]	3.462750e-20	7.791272e-15	5.432900e-13	3.486008e-71	6.312047e-26	1.000000e+00
## [32,]	1.506657e-03	9.984924e-01	1.527995e-13	1.368764e-63	9.217592e-07	2.681427e-17
## [33,]	3.070837e-07	9.999997e-01	2.184234e-11	4.624412e-82	1.480771e-09	1.951083e-08
## [34,]	2.122906e-01	7.861548e-01	1.400032e-06	6.354123e-72	1.553263e-03	1.703740e-13

Gambar 7. Nilai Probabilitas Tiap Provinsi jika masuk ke *Cluster* ke-i

Interprestasi output Berdasarkan output pada Gambar, Dengan k = 6 berdasarkan nilai probabilitas terbesarnya diperoleh (1) Provinsi: 1, 2, 4, 5, 6, 10, 12, 19, 20, 24, 25, 27, dan 29

berada pada cluster 1 ada sebanyak 13 provinsi, (2) Provinsi: 3, 16, 17, 18, 23, 28, 30, 32, 33, dan 34 ada di cluster 2 memiliki anggota sebanyak 10 provinsi, (3) Provinsi: 7, 9, 14, dan 22 ada di cluster 3 memiliki 4 provinsi, (4) Provinsi: yaitu provinsi 8 ada di cluster 4, (5) Provinsi: 11, 13, 15, dan 26 ada di cluster 5, ada 4 provinsi, dan (6) Provinsi: 21 dan 31 ada di cluster 6.

Dapat dilihat juga untuk plot dua dimensi sebagai berikut:



Gambar 8.Plot dua dimensi Cluster dan keanggotaannya

Selanjutnya adalah memberi ciri spesifik untuk menggambarkan keanggotaan masing-masing cluster yang telah terbentuk dengan melihat rata-rata masing-masing variabel (indikator kesehatan masyarakat) pada tiap masing provinsi. Untuk nilai rata-rata masing-masing variabel pada provinsi, lebih lengkapnya dapat terlihat pada Tabel 3.

Tabel 3 Nilai Rata-Rata Masing-Masing Variabel

Provinsi	Rata-Rata	Cluster
Aceh	0.028063626	Cluster 1 (Sedang)
Bali	-0.015963007	
Babel	-0.137698933	
Bengkulu	-0.009232486	
DIY	0.130475217	
Jateng	-0.098834883	
Kalbar	-0.018739495	
Sumsel	-0.000318781	
Sumbar	-0.006669113	
Sulsel	0.00142111	
Sulteng	0.009337267	
Riau	-0.065075219	
Maluku	0.048184175	
Banten	0.130634436	
Kaltara	0.303259776	
Kep Riau	0.157854936	

NTB	0.08698399	
Sultra	0.008883485	
Malut	-0.00361679	Cluster 2
Papbar	-0.00470429	(Sangat Rendah)
Sulbar	0.132603511	
NTT	-0.035866967	
Gorontalo	0.02550061	
Jakarta	-0.097175212	
Jabar	-0.065810602	Cluster 3
Kalteng	-0.141420078	(Tinggi)
Sumut	-0.30728019	
Jambi	-0.011339812	Cluster 4
		(Sangat Tinggi)
Jatim	0.039409931	
Kaltim	-0.009623838	Cluster 5
Kalsel	0.039765458	(Rendah)
Lampung	-0.071247094	
Sulut	0.053243336	Cluster 6
Papua	0.028063626	(Cukup Rendah)

Berdasarkan Tabel 3 diperoleh hasil sebagai berikut:

Cluster 1

Pada *cluster* pertama menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar -0.014410979. Oleh karena itu *cluster* pertama termasuk daerah dengan potensi kasus penyebaran Covid-19 sedang dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Aceh, Bali, Bangka Belitung, Bengkulu, DIY, Jateng, Kalimantan Barat, Sumatera Selatan, Sumatera Barat, Sulawesi Selatan, Sulawesi Tengah, Riau, Maluku.

Cluster 2

Pada *cluster* kedua menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar 0.082421626. Oleh karena itu *cluster* kedua termasuk daerah dengan potensi kasus penyebaran Covid-19 sangat rendah dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Banten, Kalimantan Utara, Kep Riau, NTB, Sulawesi Tenggara, Maluku Utara, Papua Barat, Sulawesi Barat, NTT, Gorontalo.

Cluster 3

Pada *cluster* ketiga menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar -0.06972632. Oleh karena itu *cluster* ketiga termasuk daerah dengan potensi kasus penyebaran Covid-19 tinggi dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Jakarta, Jawa barat, Kalimantan Tengah, Sumatera Utara.

Cluster 4

Pada *cluster* keempat menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar -0.30728019. Oleh karena itu *cluster* keempat termasuk daerah dengan potensi kasus penyebaran

Covid-19 sangat tinggi dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Jambi.

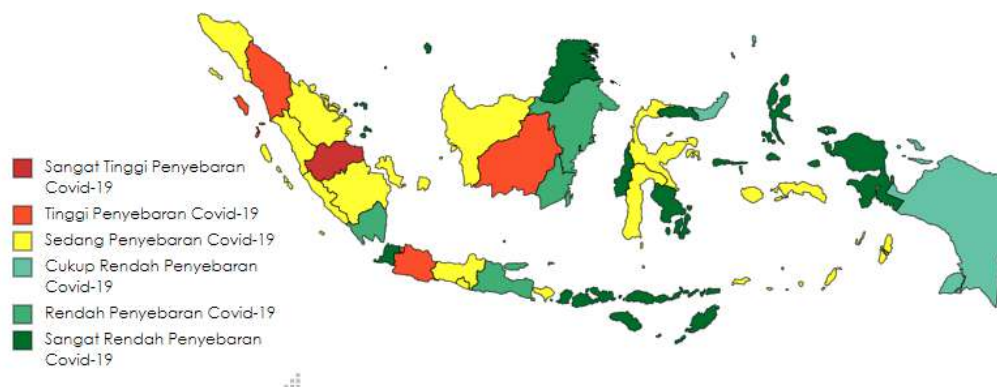
Cluster 5

Pada *cluster* kelima menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar 0.014552935. Oleh karena itu *cluster* kelima termasuk daerah dengan potensi kasus penyebaran Covid-19 rendah dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Jawa Timur, Kalimantan Timur, Kalimantan Selatan, Lampung.

Cluster 6

Pada *cluster* keenam menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar - 0.009001879. Oleh karena itu *cluster* keenam termasuk daerah dengan potensi kasus penyebaran Covid-19 cukup rendah dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Sulawesi Utara dan Papua.

Setelah melakukan *clusterisasi* daerah penyebaran Covid-19 pada 34 provinsi Indonesia, akan dilakukan penggambaran peta 2 dimensi untuk memberikan gambaran mengenai hasil *cluster* masing-masing provinsi sebagai berikut:



Dengan mengetahui pengelompokan daerah dalam penyebaran kasus Covid-19, penelitian ini diharapkan dapat menjadi masukan kepada pemerintah dalam membuat kebijakan dan dapat memberikan solusi berdasarkan skala prioritas tingkat keparahan penyebaran covid-19 di tiap provinsi secara optimal, memberikan informasi tentang tingkat keparahan penyebaran covid-19 kepada masyarakat di tiap provinsi sehingga masyarakat dapat menyikapinya dengan lebih bijaksana.

Simpulan dan Saran

Berdasarkan uraian hasil dan pembahasan di atas diperoleh 6 *cluster* provinsi di Indonesia yaitu mulai dari *Cluster* sangat tinggi, tinggi, sedang, cukup rendah, rendah, sangat rendah dalam penyebaran covid-19 yang terbentuk dengan algoritma *Gaussian Mixture Model*. Adapun secara

lengkap ke 6 cluster yang terbentuk untuk daerah penyebaran Covid-19 pada 34 provinsi Indonesia adalah sebagai berikut (1) *Cluster* 4 (sangat tinggi) yang ditempati Jambi, (2) *Cluster* 3 (tinggi) yang diisi oleh Jakarta, Jawa Barat, Kalimantan Tengah, dan Sumatera Utara, (3) *Cluster* 1 (sedang) yang ditempati Aceh, Bali, Bangka Belitung, Bengkulu, Jogjakarta, Jawa Tengah, Kalimantan Barat, Sumatera Barat, Sumatera Selatan, Sulawesi Selatan, Riau, Maluku, dan Sulawesi Tengah, (4) *Cluster* 6 (cukup rendah) yang ditempati Sulawesi Utara dan Papua, (5) *Cluster* 5 (rendah) yang ditempati Jawa Timur, Kalimantan Timur, Kalimantan Selatan, dan Lampung, (6) *Cluster* 2 (sangat rendah) yang ditempati Banten, Kalimantan Utara, Kepulauan Riau, Nusa Tenggara Barat, Sulawesi Tenggara, Maluku Utara, Papua Barat, Sulawesi Barat, Nusa Tenggara Timur, dan Gorontalo.

Berdasarkan hasil *clustering* di atas ada baiknya agar program vaksinasi dapat lebih diutamakan di beberapa provinsi berdasarkan status penyebaran Covid-19, sehingga mewabahnya virus Covid-19 dapat ditekan secara optimal.

Referensi

- Anuraga, G. (2015). Hierarchical Clustering Multiscale Bootstrap untuk Pengelompokkan Kemiskinan di Jawa Timur. *Statistika*, 1, 27–33.
- Amaluddin, Fitroh, M. Aziz Muslim, A. N. (2017). Klasifikasi Kendaraan Menggunakan Gaussian Mixture Model (GMM) dan Fuzzy Cluster Means (FCM), *Vol. 9, No(1)*, 19–24.
- Fraley, C., Raftery, A. E., Murphy, T. B., & Scrucca, L. (2012). *mclust Version 4 for R MANUAL*.
- Gio, P. U., & Irawan, D. E. (2016). Belajar Statistika dengan R. *USU Press*, 262.
- Olivia, S., Gibson, J., & Nasrudin, R. (2020). Indonesia in the Time of Covid-19. *Bulletin of Indonesian Economic Studies*. <https://doi.org/10.1080/00074918.2020.1798581>
- Pandie, E. S. Y. (2018). Implementasi Algoritma Data Mining Naive Bayes Pada Koperasi. *J-Icon*, 6(1), 15–20.
- Putra, D. K., Iwut, I., & Atmaja, R. D. (2017). Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (mfcc) Dan Gaussian Mixture Model (gmm). *eProceedings of Engineering*, 4(2), 1766–1772. Diambil dari <http://libraryeproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/487/460>
- Tarigan, A.P. (2020). *Penyakit Menular dan Virus Corona*.
- Sidabutar, J. (2020). Penerapan Metode Discrete Wavelet Transform (DWT) dan Gaussian Mixture Model (GMM) Sebagai Pengenal Penutur. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 05, 1–12. Diambil dari <http://ejournal.ust.ac.id/index.php/JTIUST/article/view/617>
- Scrucca, L. (2014). Graphical tools for model-based mixture discriminant analysis. *Advances in Data Analysis and Classification*, 8(2), 147–165. <https://doi.org/10.1007/s11634-013-0147-1>

Website KawalCovid19 (<https://kawalcovid19.id/> . Diakses tgl 15 Maret 2020 hingga 29 Oktober 2021).

Submissions

1699 / **Joko Riyono et al.** / Simulasi Clustering Provinsi di Indonesia dalam Penyebaran Covid-19 Berdas...

Library

Workflow

Publication

Submission

Review

Copyediting

Production

Round 1

Round 1 Status

Submission accepted.

Notifications

[\[majamath\] Editor Decision](#)

2022-03-20 10:58 PM

[\[majamath\] Editor Decision](#)

2022-03-20 11:19 PM

[\[majamath\] Editor Decision](#)

2022-03-20 11:20 PM

[\[majamath\] Editor Decision](#)

2022-03-24 05:05 PM

Reviewer's Attachments

Search

No Files

Revisions		Search	Upload File
▶	4833-11 Article Text, revisi manuscript Simulasi Clustering Covid-19 (1).docx (11)	March 21, 2022	Article Text
▶	4834-1 Article Text, hasil cek plagiasi sdr joko.pdf	March 20, 2022	Article Text

Review Discussions				Add discussion
Name	From	Last Reply	Replies	Closed
No Items				

Platform &
workflow by
OJS / PKP

SIMULASI CLUSTERING PROVINSI DI INDONESIA DALAM PENYEBARAN COVID-19 BERDASARKAN INDIKATOR KESEHATAN MASYARAKAT MENGUNAKAN ALGORITMA GAUSSIAN MIXTURE MODEL

by Joko Riyono FTI

Submission date: 06-Nov-2025 02:55PM (UTC+0700)

Submission ID: 2805224258

File name: 14._Majamath.pdf (3.67M)

Word count: 4676

Character count: 28234

SIMULASI *CLUSTERING* PROVINSI DI INDONESIA DALAM PENYEBARAN COVID-19 BERDASARKAN INDIKATOR KESEHATAN MASYARAKAT MENGUNAKAN ALGORITMA *GAUSSIAN MIXTURE MODEL*

Joko Riyono¹, Sofia Debi Puspa², Christina Eni Pujiastuti³
^{1,2,3}Universitas Trisakti, Jl.Kyai Tapa No.1 Grogol, Jakarta
¹jokoriyono@trisakti.ac.id

Abstrak

Gaussian Mixture Model adalah suatu metode yang mengkonstruksikan *Clustering* suatu dataset menjadi beberapa kelompok data yang memiliki distribusi Gaussian atau Normal. Pada penelitian ini akan dibahas gagasan untuk menentukan *Clustering* penyebaran Covid-19 pada 34 provinsi di Indonesia menggunakan *Gaussian Mixture Model* berdasarkan nilai indikator kesehatan masyarakat. Mengingat masih berlangsungnya pandemi Covid-19 di beberapa provinsi di Indonesia hingga saat ini, penelitian ini dipilih dengan tujuan sebagai masukan kepada pemerintah selaku pembuat kebijakan untuk bahan acuan penanganan pandemi Covid-19 sehingga program-program pencegahan penyebaran Covid-19 di tiap provinsi dapat tertangani secara lebih optimal. Hasil analisis data diperoleh 6 cluster optimal yaitu cluster yang berpotensi sangat tinggi berisi 1 provinsi, tinggi 4 provinsi, sedang 13 provinsi, cukup rendah 4 provinsi, rendah 2 provinsi, sangat rendah 10 provinsi dalam penyebaran Covid-19.

Kata Kunci: *Gaussian Mixture Model*; *Clustering*; Indikator Kesehatan Masyarakat

Abstract

An *Gaussian Mixture Model* is a method that constructs clustering a dataset into several data groups that have a Gaussian or Normal distribution. This study will discuss the idea of determining the clustering of the spread of Covid-19 in 34 provinces in Indonesia using the *Gaussian Mixture Model* based on the value of public health indicators. Considering the ongoing Covid-19 pandemic in several provinces in Indonesia to date, this research was chosen with the aim of being an input to the government as a policy maker for reference material for handling the Covid-19 pandemic so that programs to prevent the spread of Covid-19 in each province can be handled more optimally. The results of data analysis obtained 6 optimal clusters, namely a very high potential cluster containing 1 province, high 4 provinces, medium 13 provinces, quite low 4 provinces, low 2 provinces, very low 10 provinces in the spread of Covid-19.

Keywords: *Gaussian Mixture Model*; *Clustering*; Public Health Indicators

Pendahuluan

Virus *SARS-CoV-2* merupakan penyebab penyakit Covid 19. Virus ini merupakan salah satu jenis *coronavirus* yang masuk ke dalam satu kelompok virus penyebab penyakit *SARS*, yang dulu pernah terjadi pada tahun 2003 dan menewaskan sekitar 774 jiwa. Menurut penelitian yang

dilakukan (Olivia, 2020) penyakit Covid-19 sendiri setidaknya 10 kali lebih menular jika dibandingkan dari penyakit *SARS*. Penderita penyakit Covid-19 dapat mengalami demam, batuk kering, dan juga kesulitan bernapas. Keluhan yang diderita bahkan dapat berkembang menjadi *pneumonia* berat, kegagalan multiorgan bahkan kematian. Dilansir dari <https://covid19.go.id/>, sejak 2 Maret 2020 hingga 8 September 2021 telah terjadi 4.147.365 total kasus penyakit Covid-19 di Indonesia dengan menewaskan 137.782 korban jiwa. Pemerintah pusat maupun daerah pun hingga saat ini berupaya menanggulangi atau mengurangi penyebaran penyakit Covid-19 dengan mencanangkan berbagai program. Menurut (Anuraga, 2015), strategi perencanaan program seperti kewajiban memakai masker, mempraktikkan aktifitas jaga jarak, memberikan program vaksinasi, dan lain-lain bisa saja berbeda untuk tiap daerah bergantung pada pola demografi epidemi yang ditemui dan terjadi.

Mengingat masih berlangsungnya pandemi Covid-19 di beberapa provinsi dan upaya pemerintah dalam memberantas Covid-19 juga masih berlangsung, membuat keinginan untuk memilih topik identifikasi *cluster* penyebaran Covid-19 di setiap provinsi di Indonesia berdasarkan indikator kesehatan masyarakat menggunakan algoritma *Gaussian Mixture Model* berbantuan *software RStudio*. Diharapkan hasil yang diperoleh dalam penelitian ini dapat memberi masukan kepada pemerintah selaku pembuat kebijakan untuk bahan acuan penanganan pandemi Covid-19 sehingga program-program pencegahan penyebaran Covid-19 di tiap provinsi dapat tertangani secara lebih optimal.

Hasil penelitian (Diah NS & Irma Y, 2020) memberikan gambaran umum mengenai tingkat keparahan serta resiko penyebaran covid-19 pada 34 provinsi di Indonesia melalui *cluster* yang terbentuk menggunakan metode *K- Means* berdasarkan variabel *Case Fatality Rate*, Total Kasus dan kepadatan penduduk. Hasil yang didapat 3 *cluster* dengan provinsi Jakarta dan Jawa Timur memiliki tingkat keparahan yang tinggi. Adanya Indikator kesehatan masyarakat yang telah ditetapkan pemerintah sebagai patokan untuk penilaian risiko penularan Covid-19 di suatu wilayah mendorong penelitian ini untuk mengambil nya sebagai variabel dalam menentukan clustering 34 provinsi di Indonesia menggunakan *Gaussian Mixture Model*.

Algoritma *Gaussian Mixture Model* merupakan metode analisis *cluster non-hierarchical* yang bekerja untuk memodelkan beberapa data dalam distribusi Gaussian dengan parameter mean dan varians tertentu (Putra, 2017). Dalam melakukan *clustering*, Algoritma *Gaussian Mixture Model* adalah salah satu jenis *soft clustering* dimana satu *data point* bisa berada pada dua atau lebih *cluster*. Menurut (Amaluddin, 2015) Algoritma *Gaussian Mixture Model* sendiri merupakan model statistik yang sangat populer dan umum digunakan. Selain itu, algoritma ini

relatif efisien untuk pengolahan objek dalam jumlah besar. Berdasarkan hasil penelitian (Amaluddin, 2015) dan alasan bahwa tidak selalu *data point* tepat berada pada satu *cluster* tetapi bisa jadi berada di dua atau lebih *cluster* serta memperhitungkan besarnya probabilitas data point berada dalam suatu *cluster*, maka algoritma *Gaussian Mixture Model* akan digunakan dalam penelitian ini.

Dalam melakukan *clustering* menggunakan *Gaussian Mixture Model*, penentuan jumlah *cluster* yang paling sesuai untuk dataset yang dianalisa merupakan masalah yang cukup kompleks. Untuk dapat tercapainya proses ini keseimbangan antara kesempurnaan suatu model dengan tepat tidaknya suatu model terhadap data yang dianalisa harus benar-benar terjaga. Dengan kata lain, model yang dipilih untuk suatu dataset haruslah cukup sempurna agar dapat mencakup semua informasi yang terkandung di dalam dataset tersebut, tetapi tidak harus selalu sempurna agar tidak mengalami *over-fit*. (Luca Scrucca, 2016) menyatakan bahwa pada *Gaussian Mixture Model*, *cluster* yang terbentuk berbentuk *ellipsoidal* dengan pusat di μ_k . Matrik *Covarians* Σ_k menentukan karakteristik geometri yaitu *shape*, *volume* dan orientasi. Karakteristik geometrik tersebut dapat dibuat beragam antar *cluster* atau dibuat sama. Interpretasi geometrik dari berbagai parameterisasi matrik *covarians* Σ_k sebaran peluang *Gauss* pada *package Mclust* di *RStudio* dapat terlihat pada Tabel 1. Adapun jumlah *cluster* optimal dapat ditentukan berdasarkan pemilihan model terbaik. (Fraley, 2012) Pendekatan yang umum dipakai dalam menentukan pemilihan model terbaik adalah menggunakan *Bayes Information Criterion* (BIC). Nilai BIC diperoleh menggunakan rumus :

$$\text{BIC} \approx 2 \log p(x/M) + \text{konstanta} \approx 2l_M(x, \hat{\theta}) - m_M \log(n) \quad (1)$$

Dimana :

$p(x/M)$ = likelihood dari data untuk model M

$l_M(x, \hat{\theta})$ = loglikelihood mixture maksimum untuk model M

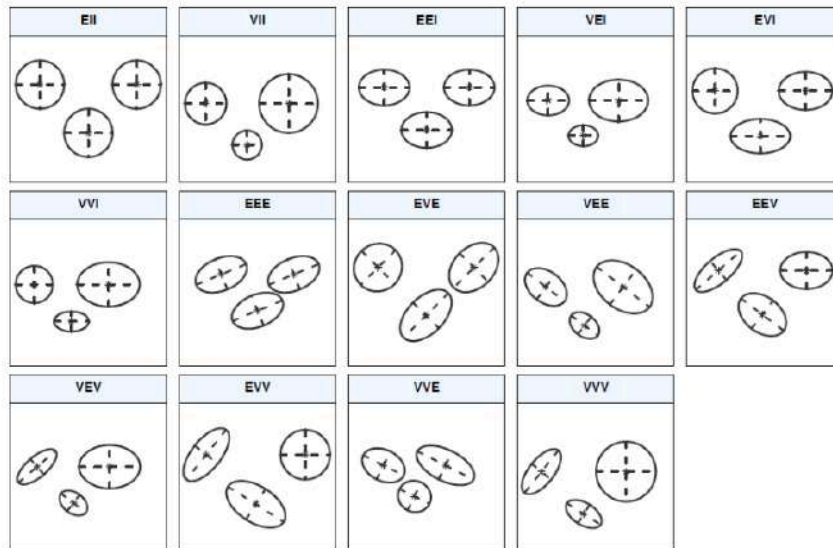
m_M = jumlah parameter bebas yang diduga dalam model

Model terbaik dipilih berdasarkan nilai BIC terbesar. Nilai BIC dapat dihitung menggunakan *RStudio* dengan *package Mclust*.

Tabel 1. Parameterisasi matriks covarians dalam grup Σ_k untuk data multidimensi tersedia dalam paket *melust*, dan karakteristik geometris yang sesuai

Model	Σ_k	Distribution	Volume	Shape	Orientation
EII	λI	Spherical	Equal	Equal	—
VII	$\lambda_k I$	Spherical	Variable	Equal	—
EEI	λA	Diagonal	Equal	Equal	Coordinate axes
VEI	$\lambda_k A$	Diagonal	Variable	Equal	Coordinate axes
EVI	λA_k	Diagonal	Equal	Variable	Coordinate axes
VVI	$\lambda_k A_k$	Diagonal	Variable	Variable	Coordinate axes
EEE	$\lambda D A D^T$	Ellipsoidal	Equal	Equal	Equal
EVE	$\lambda D A_k D^T$	Ellipsoidal	Equal	Variable	Equal
VEE	$\lambda_k D A D^T$	Ellipsoidal	Variable	Equal	Equal
VVE	$\lambda_k D A_k D^T$	Ellipsoidal	Variable	Variable	Equal
EEV	$\lambda D_k A D_k^T$	Ellipsoidal	Equal	Equal	Variable
VEV	$\lambda_k D_k A D_k^T$	Ellipsoidal	Variable	Equal	Variable
EVV	$\lambda D_k A_k D_k^T$	Ellipsoidal	Equal	Variable	Variable
VVV	$\lambda_k D_k A_k D_k^T$	Ellipsoidal	Variable	Variable	Variable

Sumber: *melust* 5 Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models by Luca Scrucca, Michael Fop, T. Brendan Murphy and Adrian E. Raftery



Gambar 1. Elips isodensitas untuk 14 model Gaussian yang diperoleh dengan dekomposisi eigen dalam kasus tiga kelompok dalam dua dimensi.

Sumber: *melust* 5 Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models by Luca Scrucca, Michael Fop, T. Brendan Murphy and Adrian E. Raftery

Gaussian Mixture Model adalah suatu metode yang memodelkan atau mengelompokkan (*clustering*) data suatu dataset menjadi beberapa kelompok data yang memiliki distribusi probabilistik Gaussian atau Normal. Metode ini mengasumsikan bahwa keseluruhan individu adalah campuran dari sebaran peluang Gaussian, mewakili distribusi Gaussian dimana masing

masing sebaran secara khas mempunyai parameter distribusi. Salah satu metode yang digunakan untuk menduga parameter adalah melalui *algoritma Expectation Maximization (EM)*. Algoritma EM merupakan algoritma iteratif populer yang dapat digunakan untuk menemukan penduga parameter dengan memaksimalkan *loglikelihood*.

EM dimulai dengan inialisasi nilai awal dengan parameter *mixture model*, kemudian secara iteratif memperbaharui bagian parameternya. Inialisasi nilai awal diperoleh menggunakan *algoritma hierarchical Clustering* sedang banyaknya cluster ditentukan menggunakan *Bayes Information Criteria (BIC)*, model terbaik dengan nilai BIC paling maksimal (Frale, 2012).

Algoritma *Expectation Maximization* adalah salah satu alternatif algoritma yang banyak digunakan dalam melakukan pemodelan *mixture*. Adapun langkah-langkah *Gaussian Mixture* dengan EM Algoritma sebagai berikut :

- 1) Inialisasi dengan m Gaussian secara acak

$$(\mu_a, \sigma_a^2), (\mu_b, \sigma_b^2), \dots, (\mu_m, \sigma_m^2)$$

- 2) Ulangi hingga konvergen

- a) E-Step apakah x_i terlihat masuk di cluster a, b, c, ..., m

$$a_i = a = P(a|x_i) = \frac{P(x_i|a) \cdot P(a)}{P(x_i)}, i = 1, 2, \dots, n$$

$$b_i = b = P(b|x_i) = \frac{P(x_i|b) \cdot P(b)}{P(x_i)}, i = 1, 2, \dots, n$$

.

.

.

$$m_i = m = P(m|x_i) = \frac{P(x_i|m) \cdot P(m)}{P(x_i)}, i = 1, 2, \dots, n \quad (2)$$

- b) M-Step perbaiki $(\mu_a, \sigma_a^2), (\mu_b, \sigma_b^2), \dots, (\mu_m, \sigma_m^2)$ sampai konvergen

$$\mu_a = \frac{a_1 x_1 + a_2 x_2 + \dots + a_n x_n}{a_1 + a_2 + \dots + a_n}$$

$$\sigma_a^2 = \frac{a_1 (x_1 - \mu_a)^2 + a_2 (x_2 - \mu_a)^2 + \dots + a_n (x_n - \mu_a)^2}{a_1 + a_2 + \dots + a_n}$$

.

.

.

$$\mu_m = \frac{m_1 x_1 + m_2 x_2 + \dots + m_n x_n}{m_1 + m_2 + \dots + m_n}$$

$$\sigma_m^2 = \frac{m_1(x_1 - \mu_a)^2 + m_2(x_2 - \mu_a)^2 + \dots + m_n(x_n - \mu_a)^2}{m_1 + m_2 + \dots + m_n} \quad (3)$$

$$P(a)_{baru} = \frac{a_1 + a_2 + \dots + a_n}{n}$$

$$P(b)_{baru} = \frac{b_1 + b_2 + \dots + b_n}{n}$$

.

.

.

$$P(m)_{baru} = \frac{m_1 + m_2 + \dots + m_n}{n} \quad (4)$$

Setelah didapat nilai $(\mu_a, \sigma_a^2), (\mu_b, \sigma_b^2), \dots, (\mu_m, \sigma_m^2)$ konvergen, hitung probabilitas cluster setiap data dari dataset berada, nilai probabilitas digunakan untuk menentukan cluster data dari dataset dengan cara melihat nilai yang terbesar.

Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah penelitian studi literatur dimana data siap pakai tentang jumlah penduduk dan data kasus covid-19 baik berupa kasus positif, sembuh dan meninggal akibat covid-19 tiap provinsi diambil dari *database* situs <https://kawalcovid19.id/> dan *database* situs resmi Kementrian Dalam Negeri Indonesia. Data yang digunakan adalah data penyebaran dari tanggal 15 Maret 2020 hingga 29 Oktober 2021. Berdasarkan data ini akan dibuat *clustering* penyebaran Covid-19 di setiap provinsi di Indonesia berdasarkan indikator kesehatan masyarakat menggunakan algoritma *Gaussian Mixture Model* berbantuan *software RStudio*. Adapun tahapan penelitian ini dilakukan sebagai berikut:

Data yang dibutuhkan dalam penelitian ini adalah data kasus Covid-19 yang diperoleh dari *database* situs <https://kawalcovid19.id/>, data jumlah penduduk yang diperoleh dari *database* situs resmi Kementrian Dalam Negeri Indonesia. Data kasus Covid-19 yang digunakan berisikan informasi berupa nilai kasus positif, sembuh, dan meninggal akibat Covid-19 per-provinsi dari 15 Maret 2020 hingga 29 Oktober 2021.

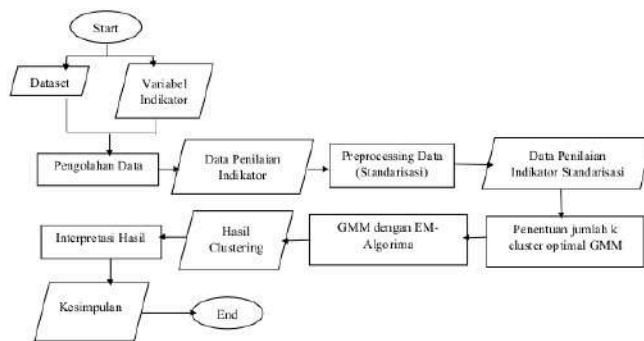
Analisis kebutuhan dilakukan dengan menentukan kebutuhan apa saja yang diperlukan dalam implementasi algoritma *Gaussian Mixture Model* untuk *clustering* daerah penyebaran Covid-19. Pada penelitian ini akan ditentukan beberapa indikator yang akan digunakan untuk menilai situasi Covid-19 di masyarakat dengan mengambil 5 indikator kesehatan masyarakat yang telah ditetapkan sebelumnya oleh pemerintah Indonesia, untuk penanganan pandemi Covid-19.

Standarisasi merupakan salah satu proses *processing* dengan menyamakan satuan dari masing-masing atribut. Dalam kasus dimana data memiliki rentang nilai yang jauh berbeda dapat

menyebabkan perhitungan menjadi tidak efektif (Pandie, 2018). Jadi standarisasi dilakukan agar data memiliki rentang nilai yang sama sehingga diharapkan data dengan nilai yang besar atau kecil tidak akan mempengaruhi hasil analisis akhir. Salah satu cara menstandarisasi data adalah dengan *Z-Score*, pada dasarnya menstandarisasi menggunakan *Z-Score* adalah mengubah nilai data asli menjadi bentuk *Z* (data berdistribusi Normal) dengan rumus :

$$z = \frac{x_i - \bar{x}}{s} \text{ dimana } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ dan } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

Perancangan penelitian merupakan rancangan langkah kerja secara keseluruhan, yang dilakukan agar dapat mempermudah implementasi. Pada penelitian ini akan dilakukan berdasarkan langkah kerja pada Gambar 2 sebagai berikut:



Gambar 2. Perancangan Sistem

Pada Gambar 2 terlihat diagram alir penelitian ini menggunakan algoritma Gaussian Mixture Model. Diawali dengan melakukan pengumpulan data, data yang digunakan pada penelitian ini adalah data kasus Covid-19 harian setiap provinsi di wilayah Indonesia dan penentuan variabel-variabel indikator kesehatan masyarakat yaitu rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif yang terjadi di tiap provinsi, rata-rata persentase penurunan angka kematian, rata-rata kenaikan jumlah sembuh dari kasus positif, rata-rata persentase penurunan jumlah meninggal dari kasus positif selama 2 minggu terakhir dari puncak kasus meninggal akibat covid-19 yang terjadi di tiap provinsi, rata-rata persentase penurunan laju insidensi kasus positif dimana dari kelima indikator tersebut semakin besar nilainya maka semakin kecil potensi penyebaran Covid-19 pada daerah tersebut. Dari sini akan dilakukan perhitungan untuk mendapatkan data nilai variabel indikator kesehatan masyarakat. Selanjutnya dilakukan preprocessing data dengan standarisasi agar data memiliki

rentang nilai yang sama sehingga diperoleh data nilai variabel indikator kesehatan masyarakat yang telah terstandarisasi. Proses selanjutnya melakukan perhitungan BIC untuk menentukan jumlah *cluster* optimal diikuti dengan analisis *Clustering* dengan *Gaussian Mixture Model* untuk mendapatkan hasil *Clustering* provinsi di Indonesia berdasarkan indikator kesehatan masyarakat. Pada langkah terakhir dilakukan penarikan kesimpulan dengan menentukan *cluster* dan provinsi yang masuk di dalam *cluster* berdasarkan nilai probabilitasnya.

Hasil dan Pembahasan

Berdasarkan data mentah Covid-19 pada 34 provinsi di Indonesia per hari mulai 15 Maret 2020 sampai 29 Oktober 2021 yaitu kasus positif, sembuh, dan meninggal akibat Covid-19 serta data jumlah penduduk akan dihitung 5 nilai variabel indikator kesehatan masyarakat yaitu: Rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif terjadi di setiap provinsi (X_1)

$$\% \text{ kasus positif} = \frac{(\text{Jumlah kasus Covid-19 pada } t - \text{Jumlah kasus Covid-19 pada } t+1)}{\text{Jumlah kasus Covid-19 pada } t} \times 100\% \quad (6)$$

Rata-rata persentase penurunan jumlah kematian dari kasus positif selama 2 minggu terakhir dari puncak kasus meninggal akibat Covid-19 terjadi di setiap provinsi (X_2)

$$\% \text{ kasus kematian} = \frac{(\text{Jumlah kasus Covid-19 pada } t - \text{Jumlah kasus Covid-19 pada } t+1)}{\text{Jumlah kasus Covid-19 pada } t} \times 100\% \quad (7)$$

Rata-rata persentase kenaikan jumlah sembuh dari kasus positif (X_3)

$$\% = \frac{(\text{Jumlah kasus sembuh Covid-19 pada } t+1 - \text{Jumlah kasus sembuh Covid-19 pada } t)}{\text{Jumlah kasus sembuh Covid-19 pada } t} \times 100\% \quad (8)$$

Rata-rata persentase penurunan laju insidensi kasus positif (X_4)

$$LI = \frac{\text{jumlah kasus positif Covid 19 pada suatu periode waktu}}{\text{jumlah waktu (person-time at risk) selama periode tersebut}} \quad (9)$$

Rata-rata persentase penurunan angka kematian (X_5)

$$CFR = \frac{\text{jumlah kasus meninggal akibat Covid 19 kumulatif}}{\text{jumlah kasus positif Covid 19 kumulatif}} \times k \quad (10)$$

Untuk $k=100.000$ artinya nilai CFR yang didapat merupakan angka kematian per 100.000 kasus positif.

Ke lima nilai variabel indikator kesehatan masyarakat yaitu : Rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif terjadi di setiap provinsi (X_1), Rata-rata persentase penurunan jumlah kematian dari kasus positif selama 2 minggu

terakhir dari puncak kasus meninggal akibat Covid-19 terjadi di setiap provinsi (X_2), Rata-rata persentase kenaikan jumlah sembuh dari kasus positif (X_3), Rata-rata persentase penurunan laju insidensi kasus positif (X_4), dan Rata-rata persentase penurunan angka kematian (X_5) di atas untuk tiap provinsi disajikan di dalam tabel 2 berikut ini:

Tabel 2 Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi

i	Provinsi	Variabel Penelitian				
		X_1	X_2	X_3	X_4	X_5
1	Aceh	-0.017395058	-0.12567459	0.47180616	-0.350422889	0.001146201
2	Bali	0.030149687	0.033427009	0.24807565	-0.173669117	0.0023349
3	Banten	0.004932763	-0.107142857	0.801298204	-0.457009101	-0.001158132
4	Bangka	-0.095474577	-0.010527865	0.146223155	-0.122875531	0.002839782
	Belitung					
5	Bengkulu	-0.076193964	-0.098095238	0.071668478	-0.589690286	0.003816346
6	DIY	0.000270674	-0.027748812	0.189858158	-0.209640353	0.001097903
	Jogjakarta					
7	Jakarta	0.08402681	-0.319930168	0.431768005	-0.072633559	0.004271964
8	Jambi	-1.336752996	-0.147340398	0.207678834	-0.256789831	-0.003196558
9	Jabar	-0.017359225	-0.675228245	0.436483966	-0.234587756	0.004815199
10	Jateng	-0.004614816	0.019290123	0.869659053	-0.234258239	0.002299963
11	Jatim	-0.025512409	0.045263981	0.113976158	-0.18890274	-0.001524052
12	Kalbar	-0.241148184	-0.257538592	0.064887385	-0.062881213	0.002506188
13	Kaltim	-0.007455939	0.011884017	0.241373218	-0.047787712	-0.000963926
14	Kalteng	0.037252511	-0.6555553	0.427742266	-0.138285453	-0.000207034
15	Kalsel	-0.029724393	0.05853192	0.291360409	-0.365975128	-0.002311997
16	Kalut	-0.022592743	-0.037221112	0.993920866	-0.281131215	0.000196384
17	Kep. Riau	0.039153866	0.081211179	1.589752179	-0.192996638	-0.000821704
18	NTB	-0.023672621	0.214285714	1.178139238	-0.580371172	0.000893519
19	Sumsel	-0.145434791	-0.168967516	0.404936954	-0.186154992	0.001922869
20	Sumbar	0.02659528	-0.060716065	0.287237516	-0.256454451	0.001743815
21	Sulut	-0.184530382	-0.771910431	1.337859852	-0.739156669	0.00150216
22	Sumut	-0.018213249	-0.543677912	0.054271858	-0.201746631	0.002265546
23	Sulteng	-0.040592164	-0.005470522	1.292578376	-0.811774395	0.000178654
24	Sulsel	-0.002795466	-0.033684119	0.306069802	-0.303767012	0.000831231
25	Sulteng	-0.051019175	0.024042525	0.3968188	-0.365563194	0.002826596
26	Lampung	0.056055904	0.053954816	0.248494695	-0.157278112	-0.002400012
27	Riau	0.018812508	0.024352882	0.135039888	-0.133431093	0.001912152
28	Maluku Utara	-0.075544727	0.058967223	0.586915878	-0.526094572	0.000173624
29	Maluku	-0.029569199	-0.222222222	0.536073196	-0.611446853	0.001788984
30	Papua Barat	-0.328061018	0.071402407	0.902740251	-0.667480196	0.003314603
31	Papua	-0.163812042	-0.749285714	1.756314314	-0.576878443	-0.000121436
32	Sulbar	-0.117340052	0.045562771	0.841715538	-0.794864684	0.001404975
33	NTT	0.068506452	-0.189044414	1.525624921	-0.742004078	-6.53253E-05
34	Gorontalo	0.021139723	-0.147363946	0.592703378	-0.646223074	0.000409086

Proses selanjutnya Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi di Tabel 2 distandarisasi. Standarisasi merupakan salah satu proses preprocessing, hasil standarisasi data Tabel 2 Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi dengan RStudio disajikan seperti berikut ini:


```
##          x1          x2          x3          x4          x5
## [1,] 0.255279028 0.03902542 -0.240589472 0.04599271 0.00077947
## [2,] 0.454011397 0.66132919 -0.706300956 0.60149509 0.70283213
## [3,] 0.348607128 0.11150971 0.444283321 -0.46059093 -1.12509072
## [4,] -0.071085698 0.48940581 -0.018122302 1.01860307 0.96703013
## [5,] 0.009505371 0.14689622 -1.075190043 -0.57671285 1.47808048
## [6,] 0.329120056 0.42204848 -0.827380853 0.64774241 0.05550472
## [7,] 0.679212481 -0.72077824 -0.324260247 1.23335337 1.71650793
## [8,] -5.259507661 -0.04571737 -0.750317669 0.44621042 -2.19180952
## [9,] 0.255428807 -2.11047724 -0.314452060 0.54110921 2.00078555
## [10,] 0.308699183 0.60603472 0.586459227 0.54251767 0.68454899
## [11,] 0.221349299 0.70762786 -0.985199141 0.73638161 -1.31657846
## [12,] -0.679907079 -0.47674228 1.007293261 1.27503802 0.79246704
## [13,] 0.296823569 0.57706677 -0.720240577 1.33955248 -1.02346126
## [14,] 0.483700502 -2.03352925 -0.332632923 0.95273613 -0.62737599
## [15,] 0.289743616 0.75952355 -0.616277937 -0.02048254 -1.72891415
## [16,] 0.233553217 0.38490889 0.844897150 0.34216755 -0.41626505
## [17,] 0.491647978 0.84823032 2.084095509 0.71888298 -0.94903596
## [18,] 0.229039431 1.36873285 1.228031859 -0.93687997 -0.05145026
## [19,] -0.279914489 -0.13030884 -0.380063000 0.74812637 0.40721405
## [20,] 0.439154322 0.29310165 -0.624852672 0.44764394 0.39351396
## [21,] -0.443330224 -2.48863615 1.560216486 -1.61558009 0.26705449
## [22,] 0.251850069 -1.59593641 -1.109371182 0.68148270 0.66653879
## [23,] 0.138317401 0.90918691 1.466040725 -1.92597153 -0.42554314
## [24,] 0.316303088 0.39803335 -0.585685555 0.24541489 -0.00404607
## [25,] 0.114733519 0.62462306 -0.396947100 -0.01072130 0.96013914
## [26,] 0.562206840 0.74162085 -0.705420431 0.87155541 -1.77497269
## [27,] 0.406623102 0.62583700 -0.941391100 0.97348522 0.48160541
## [28,] 0.012219119 0.76122618 -0.001585945 -0.70488458 -0.42817547
## [29,] 0.204392313 -0.35860721 -0.107527819 -1.06970739 0.41715125
## [30,] -1.043274193 0.00986467 0.055261020 -1.30921180 1.21551547
## [31,] -0.356725596 -2.40014272 2.430512030 -0.52195092 -0.56250165

## [32,] -0.162481238 0.70879653 0.528342710 -1.85569400 0.21619752
## [33,] 0.014338910 -0.20603603 1.950727504 -1.62775002 -0.55321005
## [34,] 0.416350631 -0.04500948 0.010450015 -1.21835211 -0.30495671
## attr(,"scaled:center")
##          x1          x2          x3          x4          x5
## -0.0784800298 -0.1356520432 0.5876784294 -0.3611831289 0.0009918373
## attr(,"scaled:scale")
##          x1          x2          x3          x4          x5
## 0.23924007 0.25566549 0.48081880 0.23395531 0.00191093
```

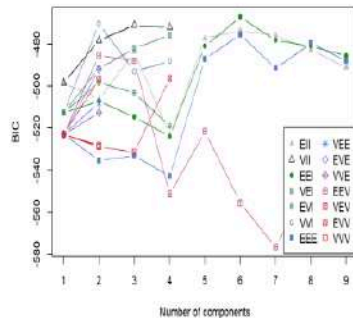
Gambar 3.Hasil standarisasi dengan RStudio Nilai Variabel Indikator Kesehatan Masyarakat Tiap Provinsi dengan RStudio

Interpretasi *output* diperoleh dari gambar 3 hasil standarisasi data nilai setiap variabel indikator kesehatan masyarakat disetiap provinsi dengan nilai *mean* dan *standard deviasinya* untuk setiap variabel seperti terangkum berikut : (1) Rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif terjadi di setiap provinsi (X_1) adalah -0,0784680298 dan 0,23924007, (2) Rata-rata persentase penurunan jumlah kematian dari kasus positif selama 2 minggu terakhir dari puncak kasus meninggal akibat Covid-19 terjadi di setiap provinsi (X_2) adalah -0,1356520432 dan 0,25566549, (3) Rata-rata persentase kenaikan jumlah sembuh dari kasus positif (X_3) adalah 0,5876784294 dan 0,48081880, (4) Rata-rata persentase penurunan laju insidensi kasus positif (X_4) adalah -0,3611831289 dan 0,23395531 dan (5) Rata-rata persentase penurunan angka kematian (X_5) adalah 0,0009918373 dan 0,00191093

Langkah berikutnya menentukan jumlah cluster dan model parameter optimal didasarkan nilai BIC, menggunakan Rstudio diperoleh:

```
## Bayesian Information Criterion (BIC):
##      EII      VII      EEI      VEI      EVI      VVI      EEV
## 1 -498.5223 -498.5223 -512.6277 -512.6277 -512.6277 -512.6277 -522.9730
## 2 -506.4091 -478.5174 -507.2429 -491.7012 -498.4362 -470.5172 -535.5590
## 3 -483.7179 -471.2150 -514.6602 -462.1353 -503.4178 -403.0061 -533.0706
## 4 -523.5063 -472.2483 -523.8824 -475.8751 -519.0186 -488.2190 -542.9600
## 5 -477.3656      NA -481.0950      NA      NA      NA -487.1784
## 6 -473.0922      NA -467.3199      NA      NA      NA -475.5705
## 7 -476.0110      NA -478.3207      NA      NA      NA -491.6081
## 8 -482.7750      NA -481.0566      NA      NA      NA -479.7325
## 9 -491.1532      NA -485.6091      NA      NA      NA -487.9913
##      VEE      EVE      VVE      EEV      VEV      EVV      VVV
## 1 -522.9730 -522.9730 -522.9730 -522.9730 -522.9730 -522.073 -522.9730
## 2 -508.0239 -492.1949 -511.7047 -485.5947 -528.8886 -496.806 -528.3212
## 3      NA      NA      NA -488.0284 -531.4230      NA      NA
## 4      NA      NA      NA -551.2766 -496.3052      NA      NA
## 5      NA      NA      NA -521.6032      NA      NA      NA
## 6      NA      NA      NA -555.5112      NA      NA      NA
## 7      NA      NA      NA -576.4611      NA      NA      NA
## 8      NA      NA      NA -545.9092      NA      NA      NA
## 9      NA      NA      NA -525.1399      NA      NA      NA
##
## Top 3 models based on the BIC criterion:
##      EEI,6      VVI,2      VII,3
## -467.3299 -470.5172 -471.2158
```

Gambar 4. Hasil Bayes Information Criterion Data Penelitian Terstandarisasi dengan Rstudio



Gambar 5. Plot Hasil Bayes Information Criterion Data Penelitian Terstandarisasi dan Jumlah Cluster dengan Rstudio

Interpretasi hasil BIC dengan package Mclust Rstudio terlihat bahwa nilai terbesar BIC sebesar -467,3299 didapat untuk Parameterisasi matriks *covarians* dalam grup matriks Covarians (Σ_k) adalah EEI dengan jumlah cluster 6. Selanjutnya akan dilakukan membuat *Cluster Gaussian Mixture Model* dengan jumlah cluster sebanyak 6 dan model parameter EEI menggunakan Rstudio diperoleh hasil sebagai berikut:

```
## -----
## Gaussian (finite mixture model) fitted by EM algorithm
## -----
## Elucent EEL (diagonal, equal volume and shape) model with 6 components:
##
## log-likelihood = dff      BIC      ZCL
##      -163.1277 -34.40 -467.3259 -470.3857
##
## Clustering table:
##      1 2 3 4 5 6
## 13 10 4 1 4 2
##
## Mixing probabilities:
##      1      2      3      4      5      6
## 0.20001593 0.20000174 0.13356540 0.02041175 0.11915753 0.05862253
##
## Means:
##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## x1 0.1397480 0.1259023 0.4071261 -0.29950100 0.5205147 -0.4000299
## x2 0.2437972 0.5261446 -1.6470652 -0.04571377 0.6081752 -0.4441804
## x3 -0.5136772 0.8250148 -0.4277380 -0.7001167 -0.7001167 1.0011642
## x4 0.5358985 -0.4270177 0.4381190 0.4421842 0.7001167 1.0011642
## x5 0.5405671 -0.1261146 0.0304158 -0.1018952 -1.4311330 0.1577436
##
## Variances:
##      [,1]
## x1 0.1025401 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1485554 0.0000000 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2341177 0.0000000 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4115583 0.0000000 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3518129 0.0000000
##
##      [,2]
## x1 0.1025401 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1485554 0.0000000 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2341177 0.0000000 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4115583 0.0000000 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3518129 0.0000000
##
##      [,3]
## x1 0.1025401 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1485554 0.0000000 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2341177 0.0000000 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4115583 0.0000000 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3518129 0.0000000
##
##      [,4]
## x1 0.1025401 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1485554 0.0000000 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2341177 0.0000000 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4115583 0.0000000 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3518129 0.0000000
##
##      [,5]
## x1 0.1025401 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1485554 0.0000000 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2341177 0.0000000 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4115583 0.0000000 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3518129 0.0000000
##
##      [,6]
## x1 0.1025401 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
## x2 0.0000000 0.1485554 0.0000000 0.0000000 0.0000000 0.0000000
## x3 0.0000000 0.0000000 0.2341177 0.0000000 0.0000000 0.0000000
## x4 0.0000000 0.0000000 0.0000000 0.4115583 0.0000000 0.0000000
## x5 0.0000000 0.0000000 0.0000000 0.0000000 0.3518129 0.0000000
```

Gambar 6. Hasil Clustering dengan GMM

Berdasarkan output di atas diperoleh hasil bahwa jumlah observasi sebanyak 34 ($n = 34$) diperoleh *clustering* GMM dengan jumlah *cluster* sebanyak 6 dan model parameter optimal adalah EEL, dimana untuk *cluster* 1 berisi 13 anggota, *cluster* 2 berisi 10 anggota, *cluster* 3 berisi 4 anggota, *cluster* 4 berisi 1 anggota, *cluster* 5 berisi 4 anggota, *cluster* 6 berisi 2 anggota. Berdasarkan output di atas diperoleh pula parameter sebaran Gaussian untuk masing-masing *cluster* sebagai berikut:

Untuk cluster 1,

$$C_1 \sim N \left(\mu_1 = \begin{pmatrix} 0.1397490 \\ 0.2437372 \\ -0.5836773 \\ 0.3393889 \\ 0.5495671 \end{pmatrix}, \sigma_1^2 = \begin{pmatrix} 0.1035401 \\ 0.1488556 \\ 0.2343173 \\ 0.4815988 \\ 0.3588329 \end{pmatrix} \right)$$

Dan seterusnya, pada cluster 6: $C_6 \sim N \left(\mu_6 = \begin{pmatrix} -0.4000299 \\ -2.4443894 \\ 1.851874201 \\ -1.535068823 \\ -0.323572286 \end{pmatrix}, \sigma_6^2 = \begin{pmatrix} 0.13061732 \\ 0.8769211 \\ 1.9953642 \\ -1.2687655 \\ -0.1577636 \end{pmatrix} \right)$

dapat diperoleh juga nilai probabilitas *cluster* setiap data dari dataset untuk menentukan cluster data dari data point dengan cara melihat nilai probabilitas terbesar, dengan hasil sebagai berikut:

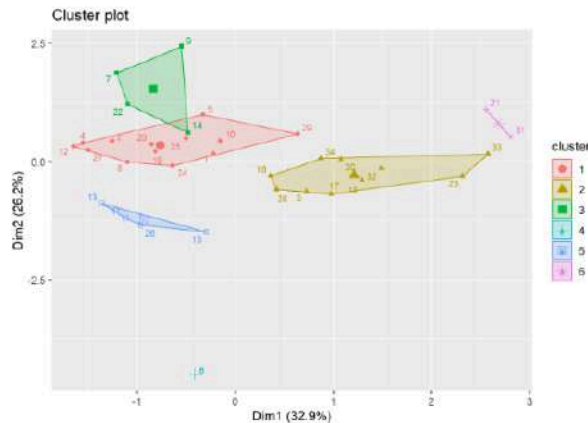
##	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
## [1,]	9.804585e-01	1.694791e-02	7.442415e-06	9.484159e-69	2.586143e-03	1.571277e-15
## [2,]	9.979986e-01	1.148552e-04	1.712992e-08	1.379142e-75	1.886542e-03	1.119737e-25
## [3,]	1.446738e-02	9.651449e-01	1.223603e-08	4.722505e-69	1.938772e-02	3.817387e-14
## [4,]	9.998498e-01	7.215096e-06	6.069754e-08	2.516135e-64	1.428880e-04	1.075828e-24
## [5,]	9.999384e-01	5.954786e-05	1.180557e-06	1.050328e-67	8.431393e-07	1.823940e-20
## [6,]	9.746883e-01	1.949595e-04	1.092151e-07	2.200377e-70	2.511867e-02	2.593171e-23
## [7,]	1.417624e-01	5.407514e-07	8.582371e-01	7.708968e-84	2.498250e-08	7.055804e-16
## [8,]	1.785151e-05	4.794725e-07	3.461265e-77	1.000000e+00	1.352383e-66	2.372094e-69
## [9,]	1.656559e-08	1.430708e-14	1.800000e+00	5.160753e-81	2.640470e-18	3.765981e-10
## [10,]	5.408674e-01	4.516788e-01	1.338073e-08	5.769333e-73	2.538043e-04	1.298365e-18
## [11,]	8.371290e-03	1.645351e-05	5.035902e-12	1.260912e-65	9.916123e-01	4.303786e-27
## [12,]	9.972012e-01	3.132144e-07	2.788717e-03	4.725632e-49	9.805278e-06	8.897969e-17
## [13,]	4.014928e-02	4.114015e-05	4.110614e-10	1.470673e-67	9.598096e-01	2.001943e-25
## [14,]	2.238252e-07	2.103030e-11	9.999996e-01	2.440776e-76	1.917167e-10	1.434418e-08
## [15,]	1.748963e-03	6.270966e-04	1.719345e-13	8.639816e-65	9.976239e-01	6.366632e-25
## [16,]	1.930858e-02	9.794087e-01	1.840910e-05	5.460168e-69	1.262688e-03	1.078230e-15
## [17,]	2.286122e-06	9.99874e-01	1.366343e-15	8.095511e-78	1.035333e-05	5.178517e-18
## [18,]	2.571785e-05	9.999731e-01	3.861337e-18	4.505327e-74	1.231873e-06	6.858899e-23
## [19,]	9.994355e-01	3.915926e-04	4.903656e-05	1.041400e-57	1.239107e-04	5.425778e-15
## [20,]	9.971147e-01	4.481276e-04	9.146827e-07	5.626932e-74	2.436226e-03	1.373048e-21
## [21,]	3.329172e-17	2.039869e-14	1.191025e-09	1.008824e-68	3.893277e-25	1.000000e+00
## [22,]	4.508793e-05	8.213376e-12	9.999541e-01	3.552717e-73	1.046972e-10	1.304698e-13
## [23,]	1.121822e-06	9.999989e-01	2.008347e-15	4.791671e-72	1.144034e-08	1.860727e-14
## [24,]	9.662613e-01	3.153352e-03	8.550362e-08	9.600726e-70	3.058526e-02	2.916629e-21
## [25,]	9.965666e-01	3.336540e-03	6.775260e-09	6.350995e-69	9.683606e-05	3.581895e-22
## [26,]	3.634320e-04	8.826697e-06	2.662998e-13	2.632589e-73	9.996277e-01	5.823714e-28
## [27,]	9.928897e-01	2.510760e-05	2.138106e-08	1.206154e-73	7.085204e-03	1.777305e-26
## [28,]	1.416374e-01	8.414002e-01	1.752640e-11	3.631820e-63	1.688237e-02	8.834618e-20
## [29,]	8.717564e-01	1.278828e-01	3.197523e-04	1.329126e-68	4.102182e-05	1.662431e-11
## [30,]	1.731589e-02	9.826841e-01	2.466900e-13	5.505252e-46	1.656255e-08	4.416188e-16
## [31,]	3.462750e-20	7.791272e-15	5.432908e-13	3.406068e-71	6.312047e-26	1.000000e+00
## [32,]	1.506657e-03	9.984924e-01	1.527995e-13	1.368764e-63	9.217592e-07	2.601427e-17
## [33,]	3.070837e-07	9.999997e-01	2.184234e-11	4.624412e-82	1.460771e-09	1.951883e-08
## [34,]	2.122906e-01	7.861548e-01	1.400832e-06	6.354123e-72	1.553263e-03	1.703740e-13

Gambar 7. Nilai Probabilitas Tiap Provinsi jika masuk ke Cluster ke-i

Interprestasi output Berdasarkan output pada Gambar, Dengan k = 6 berdasarkan nilai probabilitas terbesarnya diperoleh (1) Provinsi: 1, 2, 4, 5, 6, 10, 12, 19, 20, 24, 25, 27, dan 29

berada pada cluster 1 ada sebanyak 13 provinsi, (2) Provinsi: 3, 16, 17, 18, 23, 28, 30, 32, 33, dan 34 ada di cluster 2 memiliki anggota sebanyak 10 provinsi, (3) Provinsi: 7, 9, 14, dan 22 ada di cluster 3 memiliki 4 provinsi, (4) Provinsi: yaitu provinsi 8 ada di cluster 4, (5) Provinsi: 11, 13, 15, dan 26 ada di cluster 5, ada 4 provinsi, dan (6) Provinsi: 21 dan 31 ada di cluster 6.

Dapat dilihat juga untuk plot dua dimensi sebagai berikut:



Gambar 8. Plot dua dimensi Cluster dan keanggotaannya

Selanjutnya adalah memberi ciri spesifik untuk menggambarkan keanggotaan masing-masing cluster yang telah terbentuk dengan melihat rata-rata masing-masing variabel (indikator kesehatan masyarakat) pada tiap masing provinsi. Untuk nilai rata-rata masing-masing variabel pada provinsi, lebih lengkapnya dapat terlihat pada Tabel 3.

Tabel 3 Nilai Rata-Rata Masing-Masing Variabel

Provinsi	Rata-Rata	Cluster
Acch	0.028063626	Cluster 1 (Sedang)
Bali	-0.015963007	
Babel	-0.137698933	
Bengkulu	-0.009232486	
DIY	0.130475217	
Jateng	-0.098834883	
Kalbar	-0.018739495	
Sumsel	-0.000318781	
Sumbar	-0.006669113	
Sulsel	0.00142111	
Sulteng	0.009337267	
Riau	-0.065075219	
Maluku	0.048184175	
Banten	0.130634436	
Kaltara	0.303259776	
Kcp Riau	0.157854936	

NTB	0.08698399	Cluster 2 (Sangat Rendah)
Sultra	0.008883485	
Malut	-0.00361679	
Papbar	-0.00470429	
Sulbar	0.132603511	
NTT	-0.035866967	
Gorontalo	0.02550061	Cluster 3 (Tinggi)
Jakarta	-0.097175212	
Jabar	-0.065810602	
Kalteng	-0.141420078	
Sumut	-0.30728019	
Jambi	-0.011339812	Cluster 4 (Sangat Tinggi)
Jatim	0.039409931	
Kaltim	-0.009623838	Cluster 5 (Rendah)
Kalsel	0.039765458	
Lampung	-0.071247094	Cluster 6 (Cukup Rendah)
Sulut	0.053243336	
Papua	0.028063626	

Berdasarkan Tabel 3 diperoleh hasil sebagai berikut:

Cluster 1

Pada *cluster* pertama menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar -0.014410979. Oleh karena itu *cluster* pertama termasuk daerah dengan potensi kasus penyebaran Covid-19 sedang dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Aceh, Bali, Bangka Belitung, Bengkulu, DIY, Jateng, Kalimantan Barat, Sumatera Selatan, Sumatera Barat, Sulawesi Selatan, Sulawesi Tengah, Riau, Maluku.

Cluster 2

Pada *cluster* kedua menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar 0.082421626. Oleh karena itu *cluster* kedua termasuk daerah dengan potensi kasus penyebaran Covid-19 sangat rendah dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Banten, Kalimantan Utara, Kep Riau, NTB, Sulawesi Tenggara, Maluku Utara, Papua Barat, Sulawesi Barat, NTT, Gorontalo.

Cluster 3

Pada *cluster* ketiga menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar -0.06972632. Oleh karena itu *cluster* ketiga termasuk daerah dengan potensi kasus penyebaran Covid-19 tinggi dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Jakarta, Jawa barat, Kalimantan Tengah, Sumatera Utara.

Cluster 4

Pada *cluster* keempat menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar -0.30728019. Oleh karena itu *cluster* keempat termasuk daerah dengan potensi kasus penyebaran

Covid-19 sangat tinggi dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Jambi.

Cluster 5

Pada *cluster* kelima menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar 0.014552935. Oleh karena itu *cluster* kelima termasuk daerah dengan potensi kasus penyebaran Covid-19 rendah dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Jawa Timur, Kalimantan Timur, Kalimantan Selatan, Lampung.

Cluster 6

Pada *cluster* keenam menunjukkan bahwa rata-rata variabel masing-masing provinsi sebesar -0.009001879. Oleh karena itu *cluster* keenam termasuk daerah dengan potensi kasus penyebaran Covid-19 cukup rendah dari enam *cluster* yang terbentuk. Provinsi yang masuk dalam *cluster* ini adalah Sulawesi Utara dan Papua.

Setelah melakukan *clusterisasi* daerah penyebaran Covid-19 pada 34 provinsi Indonesia, akan dilakukan penggambaran peta 2 dimensi untuk memberikan gambaran mengenai hasil *cluster* masing-masing provinsi sebagai berikut:



Dengan mengetahui pengelompokan daerah dalam penyebaran kasus Covid-19, penelitian ini diharapkan dapat menjadi masukan kepada pemerintah dalam membuat kebijakan dan dapat memberikan solusi berdasarkan skala prioritas tingkat keparahan penyebaran covid-19 di tiap provinsi secara optimal, memberikan informasi tentang tingkat keparahan penyebaran covid-19 kepada masyarakat di tiap provinsi sehingga masyarakat dapat menyikapinya dengan lebih bijaksana.

Simpulan dan Saran

Berdasarkan uraian hasil dan pembahasan di atas diperoleh 6 *cluster* provinsi di Indonesia yaitu mulai dari *Cluster* sangat tinggi, tinggi, sedang, cukup rendah, rendah, sangat rendah dalam penyebaran covid-19 yang terbentuk dengan algoritma *Gaussian Mixture Model*. Adapun secara

lengkap ke 6 cluster yang terbentuk untuk daerah penyebaran Covid-19 pada 34 provinsi Indonesia adalah sebagai berikut (1) *Cluster* 4 (sangat tinggi) yang ditempati Jambi, (2) *Cluster* 3 (tinggi) yang diisi oleh Jakarta, Jawa Barat, Kalimantan Tengah, dan Sumatera Utara, (3) *Cluster* 1 (sedang) yang ditempati Aceh, Bali, Bangka Belitung, Bengkulu, Jogjakarta, Jawa Tengah, Kalimantan Barat, Sumatera Barat, Sumatera Selatan, Sulawesi Selatan, Riau, Maluku, dan Sulawesi Tengah, (4) *Cluster* 6 (cukup rendah) yang ditempati Sulawesi Utara dan Papua, (5) *Cluster* 5 (rendah) yang ditempati Jawa Timur, Kalimantan Timur, Kalimantan Selatan, dan Lampung, (6) *Cluster* 2 (sangat rendah) yang ditempati Banten, Kalimantan Utara, Kepulauan Riau, Nusa Tenggara Barat, Sulawesi Tenggara, Maluku Utara, Papua Barat, Sulawesi Barat, Nusa Tenggara Timur, dan Gorontalo.

Berdasarkan hasil *clustering* di atas ada baiknya agar program vaksinasi dapat lebih diutamakan di beberapa provinsi berdasarkan status penyebaran Covid-19, sehingga mewabahnya virus Covid-19 dapat ditekan secara optimal.

Referensi

- Anuraga, G. (2015). Hierarchical Clustering Multiscale Bootstrap untuk Pengelompokkan Kemiskinan di Jawa Timur. *Statistika*, 1, 27–33.
- Amaluddin, Fitroh, M. Aziz Muslim, A. N. (2017). Klasifikasi Kendaraan Menggunakan Gaussian Mixture Model (GMM) dan Fuzzy Cluster Means (FCM), *Vol. 9, No(1)*, 19–24.
- Fraley, C., Raftery, A. E., Murphy, T. B., & Scrucca, L. (2012). *mclust Version 4 for R MANUAL*.
- Gio, P. U., & Irawan, D. E. (2016). Belajar Statistika dengan R. *USU Press*, 262.
- Olivia, S., Gibson, J., & Nasrudin, R. (2020). Indonesia in the Time of Covid-19. *Bulletin of Indonesian Economic Studies*. <https://doi.org/10.1080/00074918.2020.1798581>
- Pandie, E. S. Y. (2018). Implementasi Algoritma Data Mining Naive Bayes Pada Koperasi. *J-Icon*, 6(1), 15–20.
- Putra, D. K., Iwut, I., & Atmaja, R. D. (2017). Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (mfcc) Dan Gaussian Mixture Model (gmm). *eProceedings of Engineering*, 4(2), 1766–1772. Diambil dari <http://librarye proceeding.telkomuniversity.ac.id/index.php/engineering/article/view/487/460>
- Tarigan, A.P. (2020). *Penyakit Menular dan Virus Corona*.
- Sidabutar, J. (2020). Penerapan Metode Discrete Wavelet Transform (DWT) dan Gaussian Mixture Model (GMM) Sebagai Pengenal Penutur. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 05, 1–12. Diambil dari <http://ejournal.ust.ac.id/index.php/JTIUST/article/view/617>
- Scrucca, L. (2014). Graphical tools for model-based mixture discriminant analysis. *Advances in Data Analysis and Classification*, 8(2), 147–165. <https://doi.org/10.1007/s11634-013-0147-1>

Website KawalCovid19 (<https://kawalcovid19.id/> . Diakses tgl 15 Maret 2020 hingga 29 Oktober 2021).

SIMULASI CLUSTERING PROVINSI DI INDONESIA DALAM PENYEBARAN COVID-19 BERDASARKAN INDIKATOR KESEHATAN MASYARAKAT MENGGUNAKAN ALGORITMA GAUSSIAN MIXTURE MODEL

ORIGINALITY REPORT

9%

SIMILARITY INDEX

9%

INTERNET SOURCES

4%

PUBLICATIONS

3%

STUDENT PAPERS

MATCHED SOURCE

20

ejurnal.its.ac.id

Internet Source

<1%

< 1%

★ ejurnal.its.ac.id

Internet Source

Exclude quotes On

Exclude bibliography On

Exclude matches < 10 words

SIMULASI CLUSTERING PROVINSI DI INDONESIA DALAM PENYEBARAN COVID-19 BERDASARKAN INDIKATOR KESEHATAN MASYARAKAT MENGGUNAKAN ALGORITMA GAUSSIAN MIXTURE MODEL

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

/100

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18