

**LAPORAN KEMAJUAN
PENELITIAN TERAPAN
UNIVERSITAS LAMPUNG**



**IMPLEMENTASI METODE CLUSTER NONHIERARKI
PADA PEMETAAN SEBARAN DATA COVID-19
DI INDONESIA**

**KATEGORI
Penelitian Terapan**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2021**

HALAMAN PENGESAHAN

Judul Penelitian	: Implementasi metode cluster nonhierarki pada pemetaan sebaran data COVID-19 di Indonesia tahun 2020
Manfaat sosial ekonomi Penanggulangan Covid-19 di Indonesia	: Menjadi bahan pertimbangan dalam mengambil kebijakan
Jenis penelitian	: Penelitian terapan
Ketua Peneliti	
a. Nama Lengkap	: Dr. Ir. Netti Herawati, M.Sc.
b. NIDN	: 0025016503
c. SINTA ID	: 6169478
d. Jabatan fungsional	: Lektor Kepala
e. Program studi	: Matematika
f. Nomo HP/email	: 081273809624/ netti.herawati@fmipa.unila.ac.id
Anggota Peneliti (1)	
a. Nama Lengkap	: Dr. Khoirin Nisa, M.Si.
b. NIDN	: 0026077401
c. SINTA ID	: 6050683
d. Jabatan fungsional	: Lektor
e. Program studi	: Matematika
f. Nomo HP/email	: 081379846402/ nisa.mahfudh@gmail.com
Anggota Peneliti (2)	
a. Nama Lengkap	: Subian Saidi, M.Si.
b. NIDN	: 0021088004
c. SINTA ID	: 6681591
d. Jabatan fungsional	: Asisten Ahli
e. Program studi	: Matematika
f. Nomo HP/email	: 081366351121/ subian.saidi@fmipa.unila.ac.id
Jumlah mahasiswa yang terlibat	: 2
Jumlah alumni yang terlibat	: 1
Jumlah staf yang terlibat	: 1
Lokasi kegiatan	: Lab. Matematika dan Statistika Terapan FMIPA Unila
Lama kegiatan	: 6 (enam) bulan
Biaya penelitian	: Rp. 35.000.000;
Sumber dana	: BLU Unila



Mengesahkan,
Dekan FMIPA Unila

Dr. Sripto Dwi Yuwono, M.Sc.

197307052000031001



Menyetujui,
Ketua LPPM Universitas Lampung

Dr. Lusnetta Afriani, D.E.A

NIP. 196505101993032008

Bandar Lampung, 3 Juli 2021

Ketua Peneliti,

Dr. Netti Herawati, M.Sc.
196501251990032001

IDENTITAS DAN URAIAN UMUM

1. Judul Penelitian : Implementasi metode cluster nonhierarki pada pemetaan sebaran data COVID-19 di Indonesia tahun 2020

2. Tim Peneliti :

No	Nama	Jabatan	Bidang Keahlian	Program Studi	Alokasi Waktu (jam/minggu)
1	Dr. Netti Herawati	Ketua	Statistika Terapan	Matematika	48
2	Dr. Khoirin Nisa	Anggota 1	Statistika	Matematika	30
3	Subian Saidi, S.Si., M.Si	Anggota 2	Matematika	Matematika	20

3. Objek Penelitian (Jenis Material yang akan diteliti dan segi penelitian):
Data covid-19 di Indonesia

4. Masa Pelaksanaan

Mulai : Bulan 27 April Tahun 2021

Berakhir : 3 Juli Tahun 2021

5. Usulan Biaya : Rp.35.000.000,00

6. Lokasi Penelitian : Lab Matematika dan Statistika Terapan FMIPA Unila

7. Kontributor mendasar dari hasil penelitian ini adalah menambah khasanah keilmuan statistika tentang efisiensi metode dalam menangani pencilan pada data kesehatan untuk dilakukan pengklasteran

8. Sasaran jurnal: scopus

DAFTAR ISI

RINGKASAN.....	1
BAB 1. PENDAHULUAN	2
1.1 Latar Belakang dan Masalah	2
1.2 Tujuan Khusus	3
1.3 Keutamaan Penelitian	3
1.4 Target Temuan Penelitian	3
BAB 2. TINJAUAN PUSTAKA	5
2.1 Coronavirus (COVID-19).....	5
2.2 Analisis Kluster	5
2.3 Ukuran Kemiripan Objek (jarak)	5
2.4 Metode Non-hirarki	6
2.4.1 Metode <i>Trimmed k-means</i>	6
2.5 <i>Silhouette Index</i>	8
BAB 3. METODE.....	9
3.1 Tempat dan Waktu Penelitian.....	9
3.2 Data Penelitian.....	9
3.3 Tahapan Penelitian.....	9
IV. HASIL DAN PEMBAHASAN	11
4.1 Deskriptif analisis data Covid-19 di Indonesia	11
4.2 Pendeteksian Pencilan	11
4.3 Analisis Kluster Menggunakan Metode <i>trimmed k-means</i>	14
DAFTAR PUSTAKA	

DAFTAR ISI

DAFTAR ISI

RINGKASAN	1
BAB 1. LATAR BELAKANG	2
1.1 Latar Belakang dan Masalah	2
1.2 Tujuan Khusus	3
1.3 Keutamaan Penelitian	3
1.4 Target Temuan Penelitian	4
BAB 2. TINJAUAN PUSTAKA	5
2.1 Coronavirus (COVID-19)	5
2.2 Analisis Klaster	5
2.3 Ukuran Kemiripan Objek (jarak)	5
2.4 Metode Non-hirarki	6
2.4.1 Metode Fuzzy C-Means (FCM)	6
2.4.2 Metode K-Medoids	7
2.5 Indeks Validitas Hasil Klaster	8
2.5.1 <i>Xie-Beni Index</i> (XBI)	8
2.5.2 <i>Silhouette Index</i>	9
BAB 3. METODE	10
3.1 Tempat dan Waktu Penelitian	10
3.2 Alat dan Bahan	10
3.3 Data Penelitian	10
3.4 Tahapan Penelitian	11
BAB 4. HASIL DAN PEMBAHASAN DAFTAR PUSTAKA	

RINGKASAN

IMPLEMENTASI METODE CLUSTER NONHIERARKI PADA PEMETAAN SEBARAN DATA COVID-19 DI INDONESIA TAHUN 2020

Coronavirus (COVID-19) adalah suatu penyakit menular yang ditemukan pada tahun 2019. Kasus virus ini ditemukan pertama kali di provinsi Wuhan dengan gejala antara lain batuk, demam, letih, sesak nafas, dan mengalami penurunan nafsu makan. Saat ini COVID-19 telah tersebar ke lebih dari 190 negara salah satunya yaitu negara Indonesia. Penyebaran COVID-19 yang merata di seluruh provinsi di Indonesia, merupakan penyebaran yang cukup cepat dan berdampak negatif. Luasnya wilayah Indonesia memungkinkan diperlukannya pengklasteran provinsi di Indonesia berdasarkan penyebaran COVID-19. Pengklasteran ini akan menghasilkan titik-titik pusat penyebaran kasus COVID-19 yang nantinya dapat dievaluasi dan dijadikan suatu informasi. Salah satu cabang dari ilmu statistika yang membahas tentang metode pengklasteran tersebut adalah analisis kluster. Analisis kluster merupakan suatu teknik analisis multivariat yang berguna untuk mengelompokkan data observasi ataupun variabel-variabel ke dalam kluster sedemikian rupa sehingga masing-masing kluster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran.

Penelitian ini memiliki tujuan melakukan pengklasteran provinsi berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Indonesia. Dari hasil tersebut terlihat bahwa metode trimmed k-means mampu memisahkan outlier pada data dan memberikan cluster yang optimal. Hasil ini konsisten dengan penelitian sebelumnya yang telah melaporkan metode pengelompokan k-means yang dipangkas kuat untuk outlier [16-17,24-25, 31]. Terkait data pandemi Covid-19 di Indonesia, metode trimmed k-means clustering membentuk 3 cluster dan memisahkan outlier dari 3 cluster tersebut sehingga membentuk cluster tersendiri. Kluster 1 terdiri dari 14 provinsi dan Kluster 2 dan 3 masing-masing terdiri dari 10 dan 6 provinsi. Di satu sisi, provinsi outlier yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur membentuk kluster tersendiri di luar 3 kluster di atas. Hal ini membuktikan bahwa metode trimmed k-means clustering bekerja sangat baik dalam mengelompokkan data yang mengandung outlier dan dapat digunakan pada jenis data yang sejenis.

Kata kunci: covid-19, kluster, *trimmed k-means*

BAB 1. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Coronavirus (COVID-19) adalah suatu penyakit menular yang ditemukan pada tahun 2019. Orang-orang yang terinfeksi virus ini akan mengalami penyakit pernapasan dari kategori ringan hingga menengah dan dapat sembuh tanpa harus ada perawatan khusus. Kasus virus ini ditemukan pertama kali di provinsi Wuhan dengan gejala antara lain batuk, demam, letih, sesak nafas, dan mengalami penurunan nafsu makan [1].

Saat ini COVID-19 telah tersebar ke lebih dari 190 negara salah satunya yaitu negara Indonesia. Penyebaran COVID-19 yang merata di seluruh provinsi di Indonesia, merupakan penyebaran yang cukup cepat dan berdampak negatif. Luasnya wilayah Indonesia memungkinkan diperlukannya pengklasteran provinsi di Indonesia berdasarkan penyebaran COVID-19. Pengklasteran ini akan menghasilkan titik-titik pusat penyebaran kasus COVID-19 yang nantinya dapat dievaluasi dan dijadikan suatu informasi. Salah satu cabang dari ilmu statistika yang membahas tentang metode pengklasteran tersebut adalah analisis kluster [2].

Analisis kluster merupakan suatu teknik analisis multivariat yang berguna untuk mengelompokkan data observasi ataupun variabel-variabel ke dalam kluster sedemikian rupa sehingga masing-masing kluster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran [3]. Analisis kluster dibagi menjadi dua metode yaitu metode hierarki (*hierarchical clustering methods*) dan metode non-hierarki (*non-hierarchical clustering methods*). Perbedaan antara kedua metode tersebut terletak pada penentuan jumlah kluster yang akan dihasilkan. Metode hierarki digunakan apabila jumlah kluster yang diinginkan belum diketahui,

sedangkan metode non-hierarki digunakan apabila jumlah kluster yang diinginkan telah ditentukan sebelumnya. Dalam penelitian ini akan digunakan metode non-hierarki. Metode non-hierarki yang akan digunakan dalam penelitian ini yaitu metode *trimmed k-means*. Kelebihan metode *trimmed k-means* terletak pada tingkat keakuratan yang tinggi dan waktu komputasi yang cepat [3,4] kelebihan metode *trimmed k-means* yaitu kekar terhadap pencilan.

Hal-hal tersebut yang kemudian menjadi dasar untuk mengelompokkan provinsi berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Indonesia. Dengan pengklasteran ini diharapkan dapat membantu pemerintah dalam pengklasteran wilayah berdasarkan zona sehingga pemerintah dapat lebih mudah dalam menangani kasus COVID -19 di Indonesia.

1.2 Tujuan Khusus

Tujuan khusus penelitian ini adalah pengklasteran provinsi di Indonesia berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Indonesia menggunakan metode *trimmed k-means*.

1.3 Keutamaan Penelitian

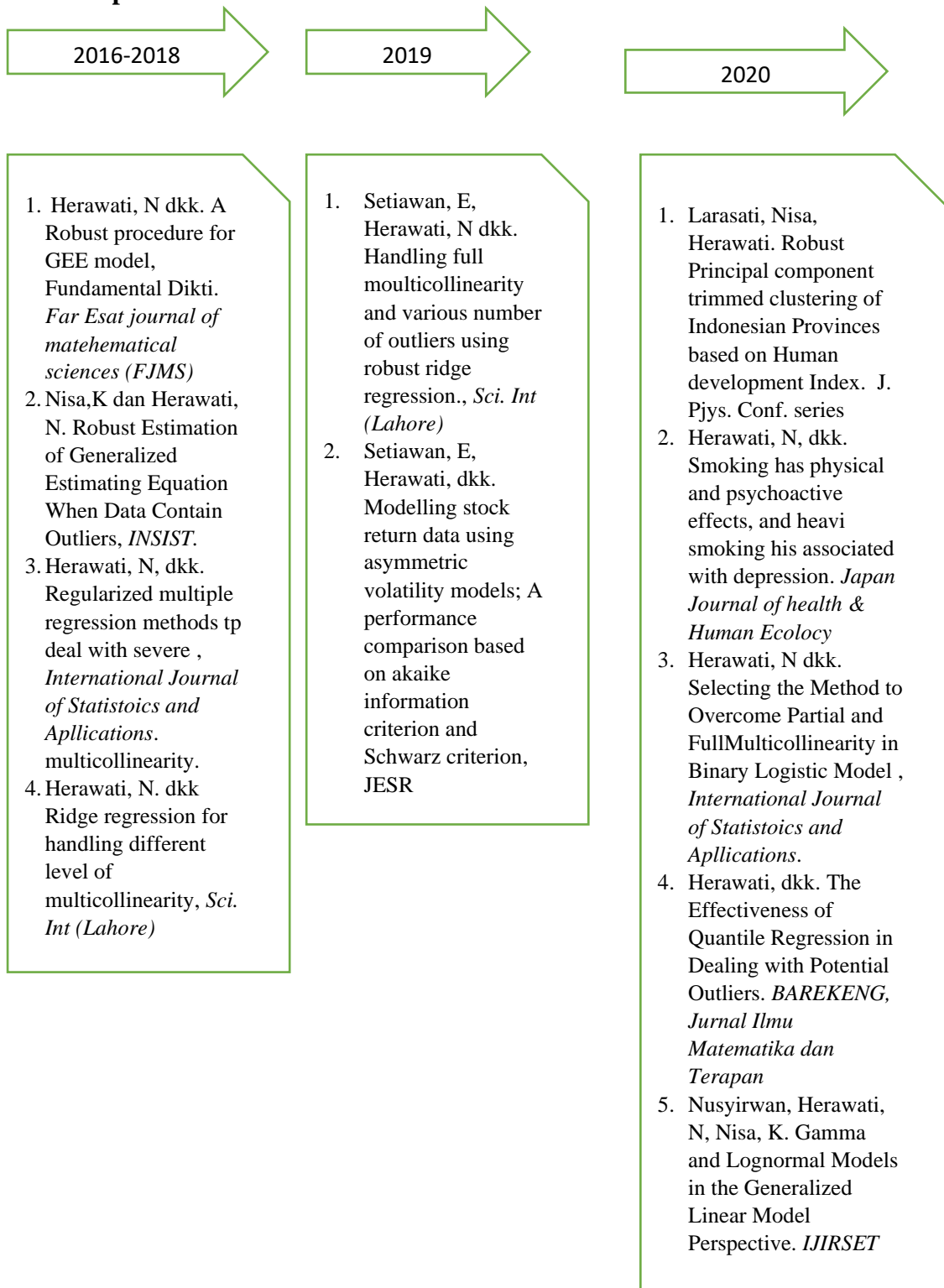
Keutamaan penelitian ini adalah sebagai berikut:

1. Mengetahui kluster provinsi di Indonesia berdasarkan zona.
2. Mengetahui perubahan kelompok provinsi pada 5 bulan pertama dan 5 bulan terakhir.
3. Membantu pemerintah dalam pengklasteran wilayah berdasarkan zona.
4. Memudahkan pemerintah dalam menangani kasus COVID-19 di Indonesia.

1.4 Target Temuan Penelitian

Target temuan utama dari penelitian ini adalah diperoleh kluster provinsi di Indonesia pada kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Indonesia dari maret 2020-Juli 2021 sehingga dapat digunakan dalam penentuan zona.

Roadmap Penelitian



BAB 2. TINJAUAN PUSTAKA

2.1 Coronavirus (COVID-19)

Coronavirus (COVID-19) adalah suatu penyakit menular yang ditemukan pada tahun 2019. Orang-orang yang terinfeksi virus ini akan mengalami penyakit pernapasan dari kategori ringan hingga menengah dan dapat sembuh tanpa harus ada perawatan khusus. Kasus virus ini ditemukan pertama kali di provinsi Wuhan dan beberapa gejala yang dialami apabila terinfeksi virus ini antara lain batuk, demam, letih, sesak nafas, dan mengalami penurunan nafsu makan [1]. Secara umum virus ini dapat menular melalui droplet atau cairan tubuh yang dikeluarkan selama bersin dan batuk [5].

2.2 Analisis Klaster

Analisis klaster merupakan suatu teknik analisis multivariat yang bertujuan untuk mengklasterkan data observasi ataupun variabel-variabel ke dalam klaster sedemikian rupa sehingga masing-masing klaster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran. Data mengenai ukuran kesamaan tersebut dapat dianalisis dengan analisis klaster sehingga dapat ditentukan siapa yang masuk klaster mana [2,3,4,6].

2.3 Ukuran Kemiripan Objek (jarak)

Dalam pengklasteran digunakan suatu ukuran yang dapat menerangkan keserupaan atau kedekatan antardata, yaitu ukuran jarak atau similaritas. Ukuran jarak yang biasa digunakan dalam analisis klaster yaitu jarak *Euclidean*. Rumus jarak *euclidean* dinyatakan sebagai berikut:

$$d_{ij} = d(x_i, x_j) = \sum_{k=1}^p \|x_{ik} - x_{jk}\| = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

dimana:

d_{ik} = jarak objek ke- i dengan objek ke- j

p = jumlah variabel

x_{ik} = data dari objek ke- i pada variabel ke- k

x_{jk} = data dari objek ke- j pada variabel ke- k

2.4 Metode Non-hirarki

Metode non-hirarki dimulai dengan menentukan terlebih dahulu jumlah kluster yang diinginkan. Setelah jumlah kluster diketahui, kemudian proses kluster dilakukan tanpa mengikuti proses hirarki [7]

2.4.1 Metode *Trimmed k-means*

K-means yang dipangkas diperkenalkan oleh [16]. Dengan menggunakan metode ini, seseorang diperbolehkan untuk menghilangkan proporsi tertentu dari kemungkinan outlier ketika mengelompokkan hasil [24,25]; proporsi yang akan dihilangkan adalah besarnya data rate yang akan dipangkas dalam metode ini. Metode trimmed k-means digunakan untuk mengatasi outlier yang terdapat pada suatu cluster data yang akan dikelompokkan. Konsep utama dari metode trimmed k-means adalah membentuk k cluster baru dengan cara menghilangkan, atau memangkas outlier yang terdapat pada data. Metode tersebut termasuk dalam kelas prosedur berdasarkan "pemangkas tidak memihak" (ditentukan oleh data) dengan tujuan membuat teknik pengelompokan hierarki klasik yaitu k-means, lebih kuat. Selanjutnya, metode k mean umum didasarkan pada minimalisasi perbedaan antara variabel acak (atau sampel variabel acak ini) dan himpunan dengan k poin yang diukur, menggunakan fungsi penalti Φ [16].

Misal α berada pada interval terbuka yaitu $\alpha \in (0,1)$, k adalah nagka real. Dan Φ adlah fungsi penalti. Untuk setiap set A pada $P(A) \geq 1-\alpha$ dan setiap k-set $M = \{m_1, m_2, \dots, m_k\}$ dalam R^p , lmaka variasi dari M pada A :

$$V_{\Phi}^A(M) = \frac{1}{P(A)} \int_A \Phi \left(\inf_{i=1, \dots, k} \|X - m_i\| \right) dP$$

Nilai $V_{\Phi}^A(M)$ mengukur berapa baik M mewakili peluang mewakili massa probabilitas dari P yang ditentukan pada A , dan tugas kita adalah memilih sekumpulan massa probabilitas yang diberikan sedemikian rupa sehingga variasinya diminimalkan. Hal ini dilakukan dengan meminimalkan $V_{\Phi}^A(M)$ pada A dan M dengan cara:

1. Cari nilai variasi- k - pada A , $V_{k,\Phi}^A(M)$, dengan meminimumkan M :

$$V_{k,\Phi}^A(M) = \inf_{M \subset R^p, \#M \neq k} V_{\Phi}^A(M);$$

2. Cari variasi trimmed k -variation, $V_{k,\Phi,\alpha}$, dengan meminimumkan A :

$$V_{k,\Phi,\alpha} := V_{k,\Phi,\alpha}(X) := V_{k,\Phi,\alpha}(P_X) := \inf_{A \in \beta^p, P(A) \geq 1-\alpha} V_{k,\Phi}^A.$$

Kita ingin trimmed set A_0 dan k -set $M_0 = \{m_1^0, m_2^0, \dots, m_k^0\}$ dengan kondisi $V_{\Phi}^{A_0}(M_0) = V_{k,\Phi,\alpha}$ [16].

Principal Component Analysis (PCA) adalah teknik statistik multivariat yang bertujuan untuk mengurangi dimensi data untuk mendapatkan variabel baru yang tidak berkorelasi dan mempertahankan sebagian besar informasi yang terkandung dalam variabel asli. Variabel yang dihasilkan merupakan kombinasi linear dari variabel asli dan disebut sebagai komponen utama (PC). Jumlah kuadrat koefisien dalam kombinasi linier sama dengan satu, dan PC-nya ortogonal.

Misal $X = (X_1, X_2, \dots, X_k)$ adalah variable acak dari normal multivariate dengan nilai tengah vector $\mu = (\mu, \mu, \dots, \mu)$ dan matrik kovarian Σ serta k independent eigenvectors, a_k , $k = 1, 2, \dots, k$. Principal component dapat ditulis sebagai::

$$Y = A(X - \mu)$$

denagn A adalah matriks $p \times p$ dari variable X pada n -element variable y . Vektor kolom A adalah Σ , i.e. $A = [a_1 | a_2 | \dots | a_k]$. Maka i -th principal component adalah:

$$y_i = a_{1i}X_1 + a_{2i}X_2 + \dots + a_{ki}X_k = a_i^T X$$

2.5 Silhoutte Index

Validasi jumlah klaster dilakukan dengan menggunakan *cluster validity index* atau indeks validitas. Indeks validitas berfungsi mengukur derajat kekompakan dan separasi struktur data pada seluruh klaster dan menemukan jumlah klaster optimal yang kompak dan terpisah dari klaster yang lain [10].

Metode validasi *silhouette index* merupakan salah satu ukuran validasi yang berbasis kriteria internal. *Silhouette index* akan mengevaluasi penempatan setiap objek dalam setiap klaster dengan membandingkan jarak rata-rata objek dalam satu klaster dan jarak antara objek dengan klaster yang berbeda [13]. Cara menghitung koefisien *silhouette* yang didefinisikan sebagai rata-rata $s(i)$ yaitu:

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \quad (2.6)$$

dengan

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}; \quad b(i) = \min d(i, C); \quad a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

Hasil perhitungan nilai koefisien *silhouette index* berada pada *range* -1 sampai 1 . Semakin besar nilai koefisien *silhouette* akan semakin baik kualitas suatu kelompok.

BAB 3. METODE

3.1 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Waktu pelaksanaan penelitian ini dilakukan dari bulan Februari sampai Juli 2021.

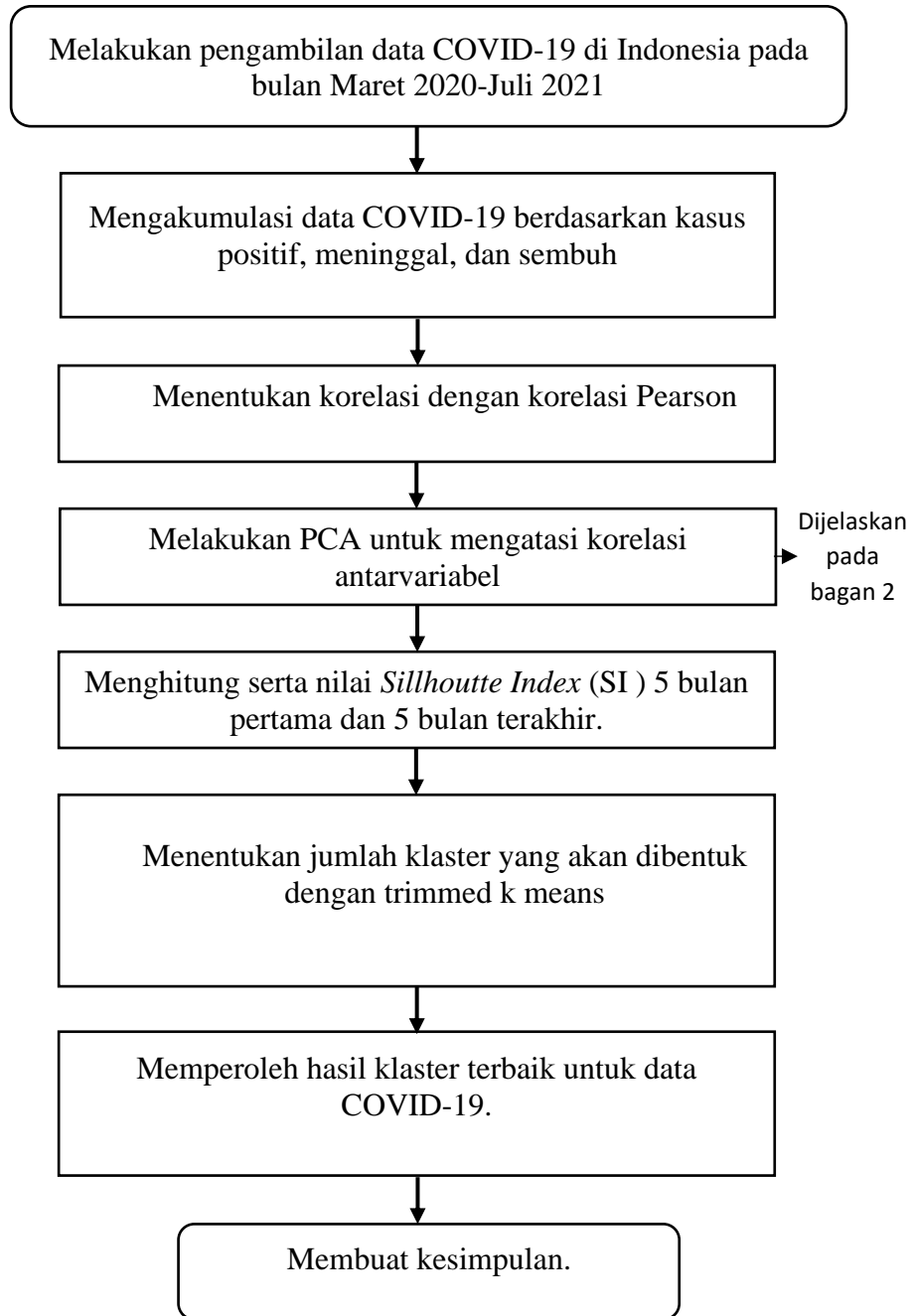
3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data COVID-19 Indonesia dari bulan Maret 2020 sd Juli 2021 yang diperoleh dari laporan analisis data COVID-19 yang disusun oleh Tim Pakar Satuan Tugas Penanganan COVID-19 dengan data bersumber dari Kementerian Kesehatan Republik Indonesia. Dengan objek 34 provinsi dan 3 variabel yaitu angka positif COVID-19, angka kematian akibat COVID-19, dan angka sembuh dari COVID-19.

3.3 Tahapan Penelitian

Tahapan penelitian ini dilakukan dengan langkah-langkah yang dijelaskan melalui bagan alir sebagai berikut

Bagan 1. Tahapan umum analisis kluster



IV.HASIL DAN PEMBAHASAN

4.1 Deskriptif analisis data Covid-19 di Indonesia

Berikut adalah hasil analisis deskriptif dari data covid-19 di 34 provinsi di Indonesia yang diambil dari 23 Maret 2020- 23 Juli 2021.

Table 1. Summary of COVID-19 cases of 34 Province in Indonesia

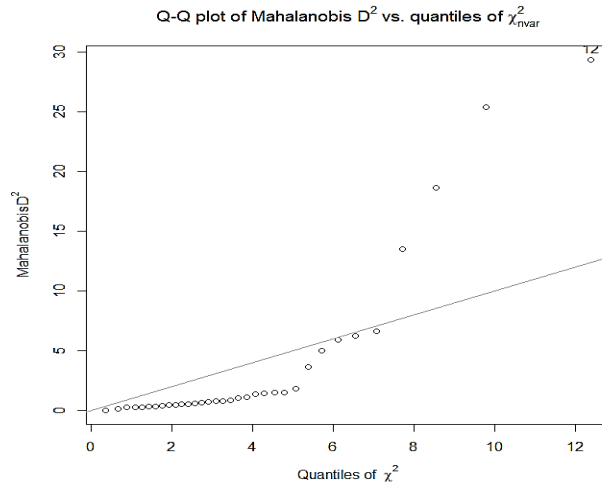
Descriptive Statistics	Confirmed Cases	Death Cases	Recovered Cases	MIR
Minimum	7103	149	5930	0.897
1 st Quartile	17790	314	13393	1.710
Median	31069	775	22580	2.250
Mean	90655	2363	71084	2.531
3 rd Quartile	73331	1706	57341	2.766
Maximum	778521	17512	678992	6.568
Standard deviation	164523.3	4292.167	136554.595	1.180

Bila kita lihat dari nilai simpangan baku yang tinggi, kita bias mengatakan bahwa kemungkinan adanya outliers pada data sangat besar. Untuk meyakinkan hal tersebut kita melakukan pendeteksian keberadaan outliers dengan menggunakan Mahalanobis squared distance method.

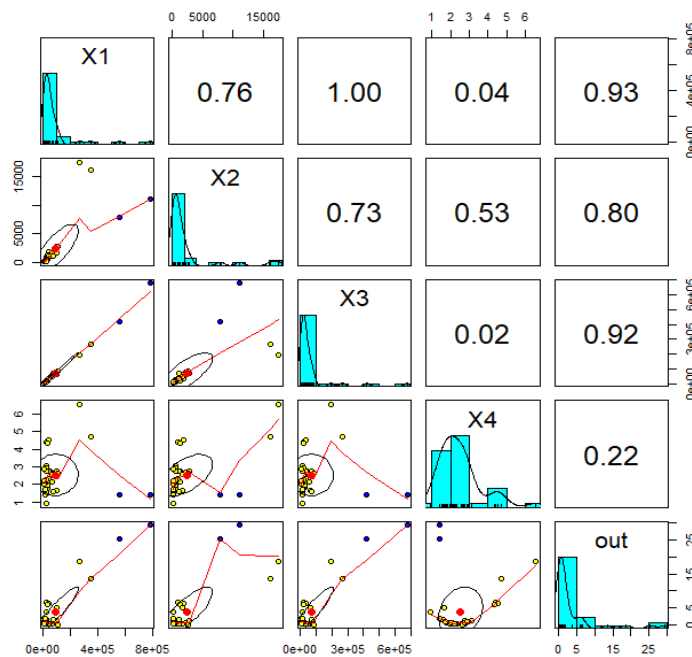
4.2 Pendeteksian Pencilan

Pendeteksian pencilan yaitu metode jarak kuadrat Mahalanobis. Pengamatan ke- i teridentifikasi pencilan jika $d_{MD}^2(i) > \chi_{p,1-\alpha}^2$. Berdasarkan kasus ini, p merupakan banyaknya variabel yang diteliti yaitu 3 variabel dan nilai α yang digunakan sebesar 5%. Maka diperoleh nilai $\chi_p^2 = 3 \cdot (1 - 0.05)$ sebesar 7,8150. Berdasarkan pendeteksian pencilan dengan membandingkan hasil jarak kuadrat Mahalanobis tiap objek dan nilai $\chi_{4,0,95}^2$, diketahui bahwa terdapat 4 provinsi yang merupakan

pencilan, yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur. Pada kasus ini data pencilan tetap disertakan dalam analisis kluster yang digunakan. Untuk Boxplot data COVID-19 di Indonesia dapat dilihat pada Gambar 1 sebagai berikut.



Gambar 1. Q-Q plot of Mahalanobis squared distance vs. chi-squared quantiles



Gambar 2. Pair panels plot of Covid-19 Data in Indonesia

Dari hasil uji Malahanonis squared distances kita dapat 4 outlier karena ada 4 titik yang ebrada diatas garis wilayah data. Berikut adalah list dari outliers dari provinsi yang mengalami penyebaran covid-19 lebih tinggi dari provinsi lain (Table 2).

Table 2. List of outliers in the Covid-19 Data in Indonesia

Province	Confirmed Cases	Death Cases	Recovered Cases	MIR
DKI				
Jakarta	778521	11131	678992	1.430
West Java	556181	7917	421977	1.423
Central Java				
Java	343210	16195	267511	4.719
East Java	266638	17512	194233	6.568

Kita juga perlu melakukan pendekteksian korelasi antarvariabel agar asumsi metode terpenuhi. Pada Tabel 3 kita lihat bahwa nilai pPearson korelasi dari variable ppositif, jumlah kematian, jumlah smebuh dan mortality incidence ratio cukup tinggi. Ini emnandakan ada korelasi anta variable.

Table 3. Correlation matrix between variables

	Confirmed Cases	Death Cases	Recovered Cases	MIR
Confirmed Cases	1.000	0.758	0.997	0.036
Death Cases	0.758	1.000	0.732	0.533
Recovered Cases	0.997	0.732	1.000	0.015
MIR	0.015	0.533	0.015	1.000

Untuk mengatasi maslaah korelasi, digunakan Principal Component Analysis sehingga korelasi antarvariabel menjadi rendah. Tabel 4 menyajikan hasil analisis dengan PCa untuk menurunkan nilai korelasi antarvariabel.

Table 4. Robust Principal Component Coefficients

Variables	PC1	PC2	PC3	PC4
Confirmed Cases	-0.561	-0.253	-0.099	0.781
Death Casess	-0.571	0.153	-0.671	-0.446
Recovered Cases	-0.558	-0.257	0.681	-0.398
MIR	-0.215	0.920	0.274	0.179

Dari Tabel 4 dapat dilihat bahwa maslaah korelasi antarvariabel telah teratasi dnegan menurunnya nilai korelasi. Selanjutnya kita bias melakukan analisis kluster dengan menggunakan trimmed k-means.

4.3 Analisis Kluster Menggunakan Metode *trimmed k-means*

Untuk menggunakan metode trimmed k-means, pertama kita cek dulu dengan silhouette index dan separation index jumlah kluster yang ada dengan menggunakan program R. Dari hasil analisis dengan Silhouette index kita dapat bahwa ada 3 cluster yang terbentuk karena nilai kurva mencapai maksimum pada $k=3$ (gambar 4)

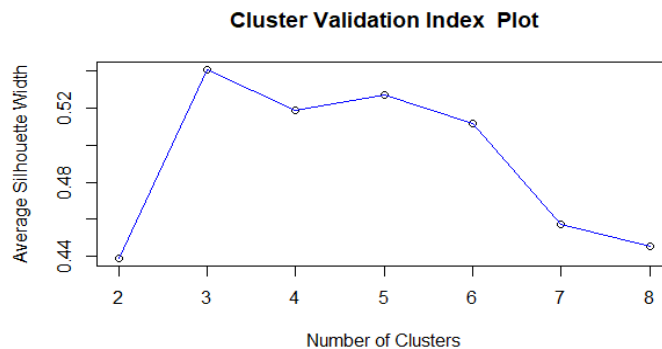


Figure 4. Plot of average silhouette index for $k=2, \dots, 8$

Selanjutnya adalah melakukan pemisahan dengan menggunakan separation index seperti tertera pada Gambar 5.

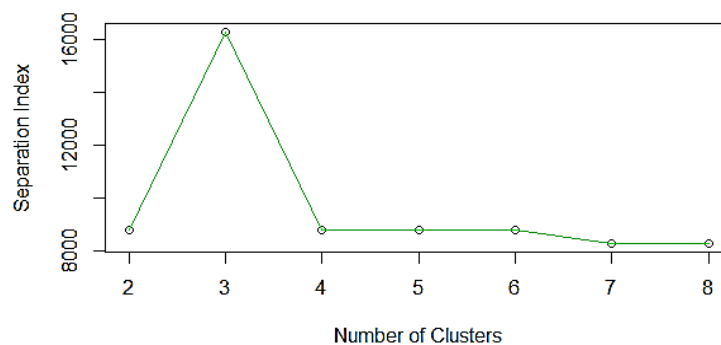


Figure 5. Plot of separation index for $k=2, \dots, 8$.

Kemudian diterapkan metode trimmed k-means dari hasil pemisahan tersebut untuk memastikan berapa jumlah kluster yang seharusnya. Hasil analisis kluster dengan trimmed k-means dapat dilihat pada Gambar 6.

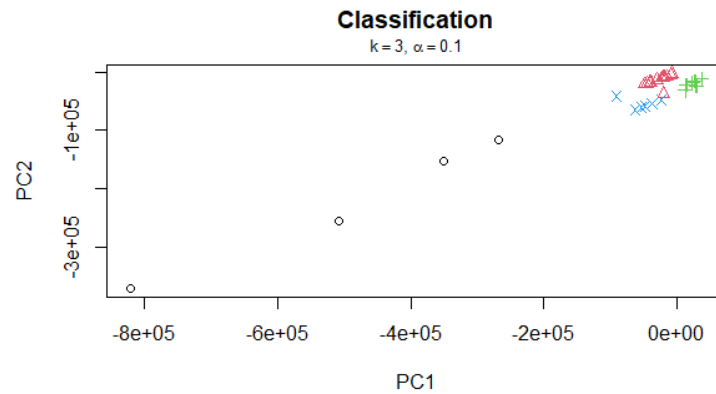


Figure 6. Trimmed *k*-means classification plot

Dari Gambar 6 dapat kita lihat bahwa trimmed *k*-means juga menghasilkan jumlah kluster sebanyak 3 ($k=3$). Hal ini sesuai dengan Silhouette indeks. Dengan menggabungkan seluruh data dengan jumlah kluster yang ada maka kita dapat 4 provinsi yang tidak bias digabungkan ke dalam ketiga kluster karena keempat provinsi ini merupakan outliers. Oleh karena itu, keempat provinsi tersebut membentuk kluster sendiri. Deskripsi jumlah kluster, provinsi yang termasuk dalam kluster dan outliers bias dilihat pada Tabel 6.

Table 6. Anggota kluster

Cluster	Cluster size	Province
1	14	North Sumatera, South Sumatera, Lampung, Riau Island, Bali, West Nusa Tenggara, West Kalimantan, South Kalimantan, North Kalimantan, Central Sulawesi, South East Sulawesi, West Sulawesi, Maluku Utara, Papua
2	10	Aceh, Jambi, Bengkulu, Bangka Belitung Island, East Nusa Tenggara, Central Kalimantan, North Sulawesi, Gorontalo, Maluku, West Papua
3	6	West Sumatera, Riau, Banten, DI Yogyakarta, East Kalimantan, South Sulawesi
outliers	4	DKI Jakarta, West Java, Central Java, and East Java

Bila kita gambarkan dalam map secara keseluruhan maka dapat dilihat pada gambar 7 berikut:

DAFTAR PUSTAKA

1. Livana, Ph.; Suwoso, R.H.; Febrianto, T.; Kushindarto, D.; Aziz, F. Dampak Pandemi Covid-19 Bagi Perekonomian Masyarakat Desa. *Indones. J. Nurs. Health Sci.* **2020**, *1*, 37–48.
2. Fahrika, A.I.; Roy, J. Dampak pandemi covid 19 terhadap perkembangan makro ekonomi di indonesia dan respon kebijakan yang ditempuh. *INOVASI* **2020**, *16*, 206–213, doi:10.29264/jinv.v16i2.8255.
3. Rohmah, S.N. Adakah Peluang Bisnis Di Tengah Kelesuan Perekonomian Akibat Pandemi Corona? *ADALAH* **2020**, *4*, 63–74.
4. Muhyiddin, M. Covid-19, New Normal, Dan Perencanaan Pembangunan Di Indonesia. *J. Perenc. Pembang. Indones. J. Dev. Plan.* **2020**, *4*, 240–252, doi:10.36574/jpp.v4i2.118.
5. Saidi, S.; Herawati, N.; Nisa, K. Modeling with Generalized Linear Model on Covid-19: Cases in Indonesia. *Int. J. Electron. Commun. Syst.* **2021**, *1*, 25–32.
6. Setiawan, S.S., Netti Herawati, Khoirin Nisa, Eri Nonparametric Modeling Using Kernel Method for the Estimation of the Covid-19 Data in Indonesia During 2020. *Int. J. Math. Trends Technol. IJMTT.*
7. Abdullah, D.; Susilo, S.; Ahmar, A.S.; Rusli, R.; Hidayat, R. The Application of K-Means Clustering for Province Clustering in Indonesia of the Risk of the COVID-19 Pandemic Based on COVID-19 Data. *Qual. Quant.* **2021**, 1–9, doi:10.1007/s11135-021-01176-w.
8. Indraputra, R.A.; Fitriana, R. K-Means Clustering Data COVID-19. *J. Tek. Ind.* **2020**, *10*, 275–282, doi:10.25105/jti.v10i3.8428.
9. Virgantari, F.; Faridhan, Y.E. K-Means Clustering of COVID-19 Cases in Indonesia's Provinces. **2020**, *7*.
10. Vahabi, N.; Salehi, M.; Duarte, J.D.; Mollalo, A.; Michailidis, G. County-Level Longitudinal Clustering of COVID-19 Mortality to Incidence Ratio in the United States. *Sci. Rep.* **2021**, *11*, 3088, doi:10.1038/s41598-021-82384-0.
11. Rojas, F.; Valenzuela, O.; Rojas, I. Estimation of COVID-19 Dynamics in the Different States of the United States Using Time-Series Clustering. *medRxiv* **2020**, 2020.06.29.20142364, doi:10.1101/2020.06.29.20142364.
12. Maugeri, A.; Barchitta, M.; Basile, G.; Agodi, A. Applying a Hierarchical Clustering on Principal Components Approach to Identify Different Patterns of the SARS-CoV-2 Epidemic across Italian Regions. *Sci. Rep.* **2021**, *11*, 7082, doi:10.1038/s41598-021-86703-3.
13. Kumar, S. Use of Cluster Analysis to Monitor Novel Coronavirus-19 Infections in Maharashtra, India. *Indian J. Med. Sci.* **2020**, *72*, 44–48, doi:10.25259/IJMS_68_2020.
14. Choi, Y.-J.; Park, M.-J.; Park, S.J.; Hong, D.; Lee, S.; Lee, K.-S.; Moon, S.; Cho, J.; Jang, Y.; Lee, D.; et al. Types of COVID-19 Clusters and Their Relationship with Social Distancing in the Seoul Metropolitan Area, South Korea. *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* **2021**, *106*, 363–369, doi:10.1016/j.ijid.2021.02.058.
15. Gallegos, M.T.; Ritter, G. A Robust Method for Cluster Analysis. *Ann. Stat.* **2005**, *33*, 347–380, doi:10.1214/009053604000000940.
16. Cuesta-Albertos, J.A.; Gordaliza, A.; Matran, C. Trimmed K-Means: An Attempt to Robustify Quantizers. *Ann. Stat.* **1997**, *25*, 553–576.
17. Garcia-Escudero, L.A.; Gordaliza, A. Robustness Properties of k Means and Trimmed k Means. *J. Am. Stat. Assoc.* **1999**, *94*, 956–969, doi:10.2307/2670010.

18. Larasati, S.D.A.; Nisa, K.; Herawati, N. Robust Principal Component Trimmed Clustering of Indonesian Provinces Based on Human Development Index Indicators. *J. Phys. Conf. Ser.* **2021**, *1751*, 012021, doi:10.1088/1742-6596/1751/1/012021.
19. Meng, S.; Fu, Y.; Liu, T.; Li, Y. Principal Component Analysis for Clustering Temporomandibular Joint Data. In Proceedings of the 2015 8th International Symposium on Computational Intelligence and Design (ISCID); December 2015; Vol. 1, pp. 422–425.
20. Rahman, A.S.; Rahman, A. Application of Principal Component Analysis and Cluster Analysis in Regional Flood Frequency Analysis: A Case Study in New South Wales, Australia. *Water* **2020**, *12*, 781, doi:10.3390/w12030781.
21. Untoro, M.C.; Anggraini, L.; Andini, M.; Retnosari, H.; Nasrulloh, M.A. Penerapan metode k-means clustering data COVID-19 di Provinsi Jakarta. *Teknol. J. Ilm. Sist. Inf.* **2021**, *11*, 59–68, doi:10.26594/teknologi.v11i2.2323.
22. Utomo, W. The Comparison of K-Means and k-Medoids Algorithms for Clustering the Spread of the Covid-19 Outbreak in Indonesia. *Ilk. J. Ilm.* **2021**, *13*, 31–35, doi:10.33096/ilkom.v13i1.763.31-35.
23. Hutagalung, J.; Ginantra, N.L.W.S.R.; Bhawika, G.W.; Parwita, W.G.S.; Wanto, A.; Panjaitan, P.D. COVID-19 Cases and Deaths in Southeast Asia Clustering Using K-Means Algorithm. *J. Phys. Conf. Ser.* **2021**, *1783*, 012027, doi:10.1088/1742-6596/1783/1/012027.
24. García-Escudero, L.; Gordaliza, A.; Matrán, C.; Mayo, A. A General Trimming Approach to Robust Cluster Analysis. *Ann. Stat.* **2008**, *36*, 1324–1345, doi:10.1214/07-AOS515.
25. García-Escudero, L.; Gordaliza, A.; Matrán, C.; Mayo, A. A Review of Robust Clustering Methods. *Adv. Data Anal. Classif.* **2010**, *4*, 89–109, doi:10.1007/s11634-010-0064-5.
26. Nisa, K.; Herawati, N.; Setiawan, E.; Nusyirwan Robust Principal Component Analysis Using Minimum Covariance Determinant Estimator.; November 30 2006.
27. Rousseeuw, P.J.; Driessen, K.V. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* **1999**, *41*, 212–223, doi:10.1080/00401706.1999.10485670.
28. Hubert, M.; Debruyne, M.; Rousseeuw, P.J. Minimum Covariance Determinant and Extensions. *WIREs Comput. Stat.* **2018**, *10*, e1421, doi:10.1002/wics.1421.
29. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65, doi:10.1016/0377-0427(87)90125-7.
30. Liu, Y.; Li, Z.; Xiong, H.; Gao, X.; Wu, J. Understanding of Internal Clustering Validation Measures. *2010 IEEE Int. Conf. Data Min.* **2010**, doi:10.1109/ICDM.2010.35.
31. Lam, B.S.Y.; Choy, S.K. A Trimmed Clustering-Based L1-Principal Component Analysis Model for Image Classification and Clustering Problems with Outliers. *Appl. Sci.* **2019**, *9*, 1562, doi:10.3390/app9081562.