

**PENERAPAN METODE *K-MEDOIDS* MENGGUNAKAN *WITHIN*
CLUSTER VARIATION DAN *GAP STATISTIK* TERHADAP DATA
COVID-19 DI INDONESIA**

SKRIPSI



TAUFIK HIDAYAT

H051191058

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR**

2023

PENERAPAN METODE *K-MEDOIDS* MENGGUNAKAN *WITHIN CLUSTER VARIATION* DAN *GAP STATISTIK* TERHADAP DATA *COVID-19* DI INDONESIA

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Sains
Pada Program Studi Statistika Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin

TAUFIK HIDAYAT
H051191058

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

JUNI 2023

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

PENERAPAN METODE *K-MEDOIDS* MENGGUNAKAN *WITHIN CLUSTER VARIATION* DAN *GAP STATISTIK* TERHADAP DATA *COVID-19* DI INDONESIA

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 20 Juni 2023



TAUFIK HIDAYAT
H051191058

PENERAPAN METODE *K-MEDOIDS* MENGGUNAKAN *WITHIN CLUSTER VARIATION* DAN *GAP STATISTIK* TERHADAP DATA *COVID-19* DI INDONESIA

Disetujui Oleh

Pembimbing Utama



Anisa, S.Si, M.Si
NIP. 197302271998022001

Pembimbing Pertama



Andi Kresna Java, S.Si, M.Si
NIP. 197312282000031001



Dr. Anna Islamiyati, S.Si, M.Si.
NIP. 197708082005012002

Pada 20 Juni 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Taufik Hidayat
NIM : H051191058
Program Studi : Statistika
Judul Skripsi : Penerapan Metode *K-Medoids* Menggunakan *Within Cluster Variaton* Dan *Gap statistik* Terhadap Data *Covid-19* Di Indonesia.

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Anisa, S.Si., M.Si (.....)
2. Sekretaris : Andi Kresna Jaya, S.Si., M.Si (.....)
3. Anggota : Dr. Erna Tri Herdiani, S.Si., M.Si (.....)
4. Anggota : Dr. Dr. Georgina Maria Tinungki, M.Si (.....)

Ditetapkan di : Makassar

Tanggal : 20 Juni 2023

KATA PENGANTAR

Segala puji senantiasa dipanjatkan kepada kehadiran Ilahirabbi Yang Mahakuasa yang telah memberikan kekuatan dan rahmat-Nya sehingga penyusunan skripsi ini dapat terselesaikan dengan baik. Skripsi dengan judul “Penerapan Metode *K-Medoids* Menggunakan *Within Cluster Variaton* Dan *Gap statistik* Terhadap Data *Covid-19* Di Indonesia” ini merupakan salah satu rangkaian syarat akademik yang harus dipenuhi untuk memperoleh Gelar Sarjana Sains pada Program Studi Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.

Skripsi ini merupakan penelitian yang bertujuan untuk membandingkan jumlah optimasi *cluster* menggunakan *within cluster variation* dan *Gap statistik* berdasarkan hasil *clusterisasi K-Medoids* terhadap data *covid-19* di Indonesia. Dalam menentukan hasil akurasi *cluster* terbaik, metode yang digunakan adalah *Average Silhouette Width*.

Penulis menyadari bahwa penyelesaian skripsi ini tidak terlepas dari bantuan dan dorongan yang diberikan oleh berbagai pihak yang secara konsisten memberikan bantuan baik secara moril maupun materil. Meskipun penulis memiliki keterbatasan dalam kemampuan dan pengetahuan, namun berkat bantuan dan dukungan tersebut, penulis berhasil menyelesaikan skripsi ini. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya dan penghargaan yang tulus kepada semua pihak yang terlibat. Oleh karena itu, dengan penuh kesadaran dan kerendahan hati, pada kesempatan ini perkenankanlah penulis menyampaikan ucapan terima kasih dan penghargaan setinggi-tingginya kepada yang terhormat:

1. Terima kasih yang tak terhingga kepada **Kedua Orang Tua** tercinta atas dukungan dan pengorbanan yang tidak bisa penulis ungkapkan dengan kata-kata terhadap proses penyelesaian skripsi ini.
2. Terima kasih kepada **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh staf jajarannya.
3. Terima kasih kepada Bapak **Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh staf jajarannya.

4. Terima kasih kepada Ibu **Dr. Anna Islamiyati, S.Si., M.Si.**, sebagai Ketua Departemen Statistika atas dorongan dan selalu menerima mahasiswa dengan ciri khasnya yang lembut.
5. Terima kasih kepada Ibu **Dr. Nurtiti Sunusi S.Si., M.Si.**, yang waktu itu masih menjabat sebagai Ketua Departemen Statistika atas dorongan dan motivasinya karena hampir setiap kali berpapasan selalu ditanya “Taufik kapan naik ujian?”. Sekali lagi terimakasih banyak, Ibu. Saat ini penulis telah menyelesaikan apa yang selalu Ibu tanyakan dengan baik.
6. Terima kasih kepada Ibu **Anisa, S.Si., M.Si.**, sebagai pembimbing utama dan penasehat akademik yang telah bersedia untuk meluangkan waktu untuk senantiasa menerima kendala penyusunan dan memberikan solusi yang terbaik.
7. Terima kasih kepada Pak **Andi Kresna Jaya, S.Si., M.Si.**, sebagai pembimbing pertama yang selalu tampil dengan ciri khas pembawaannya yang bercanda santai dengan kalimat legendnya ketika ditanyakan sesuatu “saya juga orang baru di sini, jadi kita belajar sama-sama” namun pada akhirnya tetap diberikan arahan dengan baik hingga tuntas. Sekali lagi terimakasih banyak, Pak.
8. Terima kasih kepada Ibu **Dr. Erna Tri Herdiani, M.Si.**, sebagai dosen penguji dan Pembimbing Akademik yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
9. Terima Kasih kepada **Dr. Dr. Georgina Maria Tinungki, M.Si.**, sebagai dosen penguji yang telah bersedia meluangkan waktunya untuk memberikan penilaian dan masukan terhadap skripsi ini.
10. Terima kasih kepada segenap jajaran **Dosen matakuliah** dan **Staf Departemen Statistika** yang telah banyak membantu, memberikan ilmu-ilmunya, serta berbagai kemudahan lainnya yang diberikan selama menempuh pendidikan sarjana di Departemen Statistika.
11. Terima kasih kepada **Kementrian Kesehatan Republik Indonesia** dan **Badan Nasional Penanggulangan Bencana** yang telah membuat situs web khusus untuk perkembangan kasus *covid-19* di Indonesia sehingga memudahkan penulis dalam menginput data.

12. Terima kasih kepada **Nenek** kami tercinta yang berjasa besar dalam perjalanan penulis sejak SD hingga sekarang.
13. Terima kasih kepada **Pak Ettriza Effendi, S.Pd.**, (guru Matematika) yang telah memberikan motivasi kepada penulis disaat SMA untuk berkuliah di Indonesia, motivasi yang selalu diingat ialah (“Buktikan kepada dunia bahwa kamu anak seorang buruh dari desa terpencil mampu belajar di salah satu Universitas ternama di Indonesia bahkan dunia, ilmu tak memandang uang nak, tapi ketekunan dan kesungguhan adalah jalannya”), dan sekarang penulis telah menyelesaikan Pendidikan itu bapak. Terima kasih pak.
14. Terima kasih kepada pak **Mursalim (Dosen Ilmu Komunikasi) dan Ibu** yang telah membantu penulis dalam memenuhi kebutuhan finansial selama di makassar, sekali lagi terima kasih pak. Semoga Allah membalas segala kebaikan bapak di akhirat nantinya.
15. Terima kasih kepada **Kurniawan** dan **Alfikri** yang selalu memberikan semangat kepada penulis dalam menyelesaikan proses perkuliahan di Universitas Hasanuddin.
16. Terima kasih kepada Kak **Muhammad Arbiansyah, M.Si.**, yang telah membimbing penulis melalui proses tarbiyah selama di Universitas Hasanuddin.
17. Terima kasih kepada **Reski Erik Sandi** dan **Agus Salim** yang telah menjadi sahabat seperjuangan selama berkuliah di Universitas Hasanuddin.
18. Terima kasih kepada warga istana yang setia menemani penulis dari masa kuliah perdana sampai akhir perkuliahan. Terima kasih kanda-kanda: **Agus Hermawan, Arief Rahman Nur, Fadilah Amirul Adhel, Muhammad Syamsul Bahri, Muhammad Fathurahman, Muhammad Ferdiansyah, dan Sapriadi Rasyid.**
19. Terima kasih kepada, **Nurazizah** dan **Muliana**, dan **Nurmawati** yang telah memberikan masukan tentang penulisan skripsi.
20. Terima kasih kepada **Ikhwah** di **LDK MPM Universitas Hasanuddin** yang telah memberikan ruang kepada saya dalam belajar dan berdakwah.
21. Terima kasih kepada seluruh elemen **Masjid Al-Aqsho Unhas** yang telah memberikan ruang kepada saya dalam membangun dan membentuk suatu

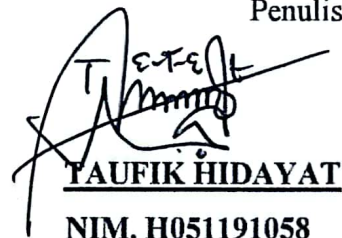
komunitas jangka Panjang yang berorientasi pada poros dakwah di Universitas Hasanuddin yakni Sahabat Aqsho.

22. Terima kasih kepada teman-teman dan adik-adik di **Sahabat Aqsho** yang menjadi teman dakwah penulis di masa-masa akhir perkuliahan, semoga komunitas yang kita bentuk dapat menjadi amal jariyah nantinya, aamin ya rabbal 'alamin.
23. Terima kasih kepada teman-teman sejawat di **BEM FMIPA Unhas dan Himastat** yang selalu memberikan pengalaman berharga dalam berorganisasi.
24. Terima kasih kepada teman-teman **Komunitas ANBISI P2JUH** yang telah menjadi teman bagi penulis di saat berkuliah di unhas sejak semester 2 hingga akhir.
25. Terima kasih kepada semua teman-teman di **Angkatan Statistika 2019** yang telah menerima penulis yang merupakan orang pendatang di Makassar.
26. Terima kasih yang setinggi-tingginya kepada **seluruh pihak** yang mungkin tidak sempat penulis sebutkan satu persatu. Terima kasih atas segala dukungan, partisipasi, dan apresiasinya yang diberikan kepada penulis.

Penulis juga menyadari bahwa skripsi ini masih jauh dari kata sempurna, namun ini hasil terbaik yang dapat diberikan oleh penulis pada penelitian ini. Oleh karena dengan segala kerendahan hati penulis mengucapkan permohonan maaf yang sebesar-besarnya. Akhir kata, semoga tulisan ini dapat memberikan manfaat untuk berbagai pihak.

Makassar, 20 Juni 2023

Penulis



TAUFIK HIDAYAT
NIM. H051191058

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR
UNTUK KEPENTINGAN AKADEMIK

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Taufik Hidayat

NIM : H051191058

Program Studi : Statistika

Departemen : Statistika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty- Free Right)** atas tugas akhir saya yang berjudul:

“Penerapan Metode *K-Medoids* Menggunakan *Within Cluster Variaton* Dan *Gap statistik* Terhadap Data *Covid-19* Di Indonesia”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar, 20 Juni 2023

Yang Menyatakan,



TAUFIK HIDAYAT

NIM. H051191058

ABSTRAK

Pada pertengahan Desember 2019, virus yang tidak diketahui mulai menyebar di pasar ikan di Wuhan, Provinsi Hubei, China. Menurut data dari situs resmi Kementerian Kesehatan Republik Indonesia, situs pemerintah tentang *covid-19*, dan Badan Nasional Penanggulangan Bencana (BNPB), virus *covid-19* terdeteksi masuk ke Indonesia pada tanggal 2 Maret 2020 di Provinsi Jawa Barat. Merebaknya kasus *covid-19* di Indonesia dalam kurun waktu 2020 – 2022 mendorong perlunya pengelompokan daerah di Indonesia menggunakan teknik *clustering*. Hal ini bertujuan untuk mengidentifikasi daerah yang memiliki tingkat kasus *covid-19* yang padat dan rendah. Dengan demikian, pemerintah dapat mengambil langkah-langkah yang sesuai dalam menangani kasus *covid-19* di Indonesia. Dalam penelitian ini, digunakan dua metode untuk menentukan jumlah *cluster* yang optimal, yaitu *within-cluster variation* dan *Gap statistik*. Teknik *clustering* yang digunakan adalah *K-Medoids*, yang mampu mengatasi adanya outlier dalam data. Tujuan penelitian ini adalah untuk memperoleh *cluster* optimal dan mengelompokkan provinsi-provinsi berdasarkan hasil analisis menggunakan metode *K-Medoids*. Berdasarkan analisis *cluster* menggunakan metode *K-Medoids* untuk *covid-19* sebelum dan setelah vaksinasi, ditemukan jumlah *cluster* optimal: 2 *cluster* dengan *within cluster variation*, 8 *cluster* sebelum vaksinasi, dan 6 *cluster* setelah vaksinasi menggunakan *gap statistik*. Evaluasi *cluster* menunjukkan bahwa metode *within cluster variation* memperoleh hasil nilai *average silhouette width* 0,85 baik sebelum dan sesudah vaksin. Sementara *gap statistik* memperoleh nilai 0,47 sebelum vaksin dan 0,56 setelah vaksin. Dengan demikian, metode *within cluster variation* dapat digunakan untuk mengoptimalkan jumlah *cluster* dalam analisis *K-Medoids* pada kasus *covid-19*.

Kata kunci: *Clustering, K-Medoids, within-cluster variation, Gap statistik, covid-19, clustering, Average Silhouette Width.*

ABSTRACT

In mid-December 2019, an unknown virus began spreading in the seafood market in Wuhan, Hubei Province, China. According to data from the official website of the Ministry of Health of the Republic of Indonesia, the government's website on Covid-19, and the National Disaster Management Agency (BNPB), the Covid-19 virus was detected in Indonesia on March 2, 2020, in West Java Province. The outbreak of Covid-19 cases in Indonesia from 2020 to 2022 has prompted the need for regional grouping in Indonesia using clustering techniques. This aims to identify areas with high and low levels of Covid-19 cases, enabling the government to take appropriate measures in handling the Covid-19 cases in Indonesia. This study utilized two methods, namely within-cluster variation and Gap statistic, to determine the optimal number of clusters. The K-Medoids clustering technique was employed, which can handle outliers in the data. The objective of this research was to obtain the optimal clusters and group the provinces based on the analysis using the K-Medoids method. Based on the cluster analysis using the K-Medoids method for Covid-19 before and after vaccination, the optimal number of clusters was found to be 2 clusters using within-cluster variation, 8 clusters before vaccination, and 6 clusters after vaccination using the Gap statistic. Cluster evaluation demonstrated that the within-cluster variation method achieved a good result with an average silhouette width of 0.85 both before and after vaccination. Meanwhile, the Gap statistic obtained a value of 0.47 before vaccination and 0.56 after vaccination. Thus, the within-cluster variation method can be used to optimize the number of clusters in the K-Medoids analysis for Covid-19 cases. These findings provide valuable insights for the management of Covid-19 in Indonesia.

Keywords: *Clustering, K-Medoids, within-cluster variation, Gap statistik, covid-19, sverage silhouette width.*

DAFTAR ISI

HALAMAN SAMPUL.....	I
HALAMAN JUDUL.....	II
LEMBAR PERNYATAAN KEOTENTIKAN.....	III
HALAMAN PENGESAHAN.....	V
KATA PENGANTAR.....	VI
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR.....	X
UNTUK KEPENTINGAN AKADEMIK.....	X
ABSTRAK.....	XI
ABSTRACT.....	XII
DAFTAR ISI.....	XIII
DAFTAR GAMBAR.....	XV
DAFTAR TABEL.....	XVI
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 <i>Clustering</i>	6
2.1.1 <i>K-Medoids</i>	7
2.1.2 Jumlah <i>cluster</i> optimal.....	8
2.1.2.1. Normalisasi Data.....	9
2.1.2.2. <i>Within cluster variation</i>	10
2.1.2.3. Gap statistik.....	11
2.1.3 <i>Average Silhouette Width</i>	11
2.2 <i>Covid-19</i>	13
BAB III METODOLOGI PENELITIAN.....	16
3.1 Sumber Data.....	16
3.2 Variabel Penelitian.....	16
3.3 Metode Analisis.....	16
BAB IV HASIL DAN PEMBAHASAN.....	18
4.1 Kasus <i>Covid-19</i> Sebelum Vaksin.....	18
4.1.1 Statistika Deskriptif.....	18
4.1.2 Normalisasi Data.....	19

4.1.3	Penentuan Jumlah <i>Cluster</i>	19
4.1.3.1	Uji <i>Within cluster variation</i>	19
4.1.3.2	Gap statistik.....	20
4.1.4	Analisis <i>-K-Medoids</i>	21
4.1.4.1	Menggunakan hasil Uji <i>Within cluster variation</i>	21
4.1.4.2	Menggunakan hasil Gap statistik.....	26
4.1.5	Interpretasi Hasil <i>Cluster</i>	33
4.1.6	Evaluasi Hasil <i>Cluster</i>	34
4.2	Kasus <i>Covid-19</i> Setelah Vaksin.....	36
4.2.1	Statistika Deskriptif.....	36
4.2.2	Normalisasi Data.....	38
4.2.3	Penentuan Jumlah <i>Cluster</i>	38
4.2.3.1	Uji <i>Within cluster variation</i>	38
4.2.3.2	Gap statistik.....	39
4.2.4	Analisis <i>-K-Medoids</i>	39
4.2.4.1	Menggunakan hasil Uji <i>Within cluster variation</i>	39
4.2.4.2	Menggunakan hasil Gap statistik.....	44
4.2.5	Interpretasi Hasil <i>Cluster</i>	49
4.2.6	Evaluasi Hasil <i>Cluster</i>	50
BAB V KESIMPULAN DAN SARAN.....		52
5.1	Kesimpulan.....	52
5.2	Saran	53
DAFTAR PUSTAKA.....		54
LAMPIRAN.....		52

DAFTAR GAMBAR

Gambar 4.1 Grafik Hasil *Within cluster variation*.....20

Gambar 4.2 Grafik Hasil Gap statistik.....20

Gambar 4.3 Hasil *Cluster Kasus Covid-19* Menggunakan Metode *K-Medoids*..33

Gambar 4.4 Perbandingan Hasil Evaluasi *Cluster Uji Within cluster variation* (A) Dan Gap statistik (B).....36

Gambar 4.5 Grafik Hasil *Within cluster variation*.....38

Gambar 4.6 Grafik Hasil Gap statistik.....39

Gambar 4.7 Hasil *Cluster Kasus Covid-19* Sesudah Vaksin Menggunakan Metode *K-Medoids*.....49

Gambar 4.8 Perbandingan Hasil Evaluasi *Cluster Uji Within cluster variation* (A)51

DAFTAR TABEL

tabel 4.1	Statistika Dekskriptif Kasus <i>Covid-19</i> Sebelum Masa Vaksin.....	18
Tabel 4.2	Hasil Normalisasi Data Menggunakan <i>Min-Max Normalization</i>	19
Tabel 4.3	<i>Medoid</i> Awal.....	21
Tabel 4.4	Hasil Perhitungan Jarak Antara <i>Medoid</i> Dengan Objek Lainnya.....	22
Tabel 4.5	<i>Non-Medoid</i> Iterasi 1.....	22
Tabel 4.6	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 1.....	23
Tabel 4.7	<i>Non-Medoid</i> Iterasi 2.....	24
Tabel 4.8	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 2.....	24
Tabel 4.9	<i>Non-Medoid</i> Iterasi 3.....	25
Tabel 4.10	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 3.....	25
Tabel 4.11	<i>Non-Medoid</i> Iterasi 4.....	25
Tabel 4.12	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 4.....	26
Tabel 4.13	<i>Medoid</i> Awal.....	26
Tabel 4.14	Hasil Perhitungan <i>Medoid</i> Dengan Antar Objek Lainnya.....	27
Tabel 4.15	<i>Non-Medoid</i> Iterasi 1.....	28
Tabel 4.16	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 1.....	28
Tabel 4.17	<i>Non-Medoid</i> Iterasi 2.....	29
Tabel 4.18	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 2.....	30
Tabel 4.19	<i>Non-Medoid</i> Iterasi 3.....	30
Tabel 4.20	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 3.....	31
Tabel 4.21	<i>Non-Medoid</i> Iterasi 4.....	32
Tabel 4.22	Hasil Perhitungan <i>Non-Medoid</i> Dengan Antar Objek Lainnya Pada Iterasi 2.....	32
Tabel 4.23	Statistik Dekskriptif Kasus <i>Covid-19</i> Sebelum Masa Vaksin.....	37
Tabel 4.24	Hasil Normalisasi Data Menggunakan <i>Min-Max Normalization</i>	38
Tabel 4.25	<i>Medoid</i> Awal.....	40

Tabel 4.26 Hasil Perhitungan Jarak Antara *Medoid* Dengan Objek Lainnya.....40

Tabel 4.27 *Non-Medoid* Iterasi 1.....41

Tabel 4.28 Hasil Perhitungan Jarak Antara *Non-Medoid* Dengan Objek Lainnya ..
.....41

Tabel 4.29 *Non-Medoid* Iterasi 2.....41

Tabel 4.30 Hasil Perhitungan *Non-Medoid* Dengan Antar Objek Lainnya Pada
Iterasi 2.....42

Tabel 4.31 *Non-Medoid* Iterasi 3.....42

Tabel 4.32 Hasil Perhitungan *Non-Medoid* Dengan Antar Objek Lainnya Pada
Iterasi 3.....43

Tabel 4.33 *Non-Medoid* Iterasi 4.....43

Tabel 4.34 Hasil Perhitungan *Non-Medoid* Dengan Antar Objek Lainnya Pada
Iterasi 4.....44

Tabel 4.35 *Medoid* Awal.....44

Tabel 4.36 Hasil Perhitungan *Medoid* Dengan Antar Objek Lainnya.....45

Tabel 4.37 *Non-Medoid* Iterasi 1.....45

Tabel 4.38 Hasil Perhitungan *Non-Medoid* Dengan Antar Objek Lainnya Pada
Iterasi 1.....46

Tabel 4.39 *Non-Medoid* Iterasi 2.....46

Tabel 4.40 Hasil Perhitungan *Non-Medoid* Dengan Antar Objek Lainnya Pada
Iterasi 2.....47

Tabel 4.41 *Non-Medoid* Iterasi 3.....47

Tabel 4.42 Hasil Perhitungan *Non-Medoid* Dengan Antar Objek Lainnya Pada
Iterasi 2.....48

Tabel 4.43 *Non-Medoid* Iterasi 4.....48

Tabel 4.44 Hasil Perhitungan *Non-Medoid* Dengan Antar Objek Lainnya Pada
Iterasi 2.....49

BAB I PENDAHULUAN

1.1 Latar Belakang

Pertengahan Desember tahun 2019, virus yang tak dikenal mulai menyebar di pasar ikan di Wuhan, Provinsi Hubei, China. Laporan menunjukkan bahwa sejak tanggal 31 Desember 2019 hingga 3 Januari 2020, kasus virus ini meningkat dengan cepat, dengan dilaporkannya sebanyak 44 kasus. *World Health Organization* (WHO) mengumumkan bahwa perkembangan virus di China akan menimbulkan virus varian baru yaitu *coronavirus* jenis baru yang dikenal sebagai SARS-CoV-2. Pada awalnya, penyakit ini diberi nama sebagai 2019 *novel coronavirus* (2019-nCoV), namun pada 11 Februari 2020, WHO mengumumkan nama baru yaitu penyakit *coronavirus (covid 19)* yang disebabkan oleh virus SARS-CoV-2.

Berdasarkan data dari situs resmi Kementerian Kesehatan Republik Indonesia, situs pemerintah tentang *Covid-19*, dan Badan Nasional Penanggulangan Bencana, telah terdeteksi bahwa virus *Covid-19* sudah masuk di Indonesia pada tanggal 2 Maret 2020 di Provinsi Jawa barat. Hingga Sabtu (31/12/2022), jumlah orang yang terkonfirmasi positif virus corona di Indonesia telah mencapai 6.719.815, serta 160.612 orang meninggal, 9.871 orang positif aktif (dirawat), dan 6.549.332 orang dinyatakan sembuh.

Dalam penyebaran *Covid-19* di Indonesia, ada beberapa daerah yang lebih rentan terhadap penyebarannya dibandingkan daerah lainnya. Oleh karena itu, penelitian ini dibagi menjadi dua periode yaitu sebelum dan setelah ditemukannya vaksin untuk mengelompokkan perbedaan dampak pandemi di daerah-daerah dengan tingkat kerawanan *Covid-19* yang berbeda di Indonesia. Untuk mengetahui pengelompokan provinsi di Indonesia menggunakan teknik *clustering*.

Clustering adalah teknik analisis data yang bertujuan untuk membagi data menjadi kelompok-kelompok yang memiliki karakteristik yang serupa atau sama. Data yang memiliki karakteristik berbeda akan dikelompokkan ke dalam *cluster* yang berbeda. *Clustering* berbeda dengan metode klasifikasi dalam menyelesaikan dataset. Dalam menentukan jumlah *cluster* yang optimal, diperlukan suatu metode

analisis yang dapat digunakan. Pada penelitian ini, metode yang digunakan untuk optimasi jumlah *cluster* adalah *within-cluster variation* dan *Gap statistik*. Metode ini digunakan untuk memperkirakan jumlah *cluster* yang paling sesuai dengan data terhadap data *Covid-19*.

Perbedaan data dalam setiap *cluster* menjadi fokus utama konsep *within cluster variation*. Tujuan utamanya adalah untuk memperkecil variasi data dalam setiap *cluster*, sehingga setiap *cluster* memiliki data yang seragam. Penurunan dalam *within cluster variation* dapat digunakan sebagai kriteria dalam menentukan jumlah *cluster* yang optimal dalam dataset tertentu. Jumlah *cluster* optimal dapat dipilih ketika penurunan dalam *within cluster variation* tidak lagi signifikan saat menambahkan lebih banyak *cluster*. Ketika menambahkan lebih banyak *cluster* tidak memberikan penurunan yang signifikan dalam *within cluster variation*, maka jumlah *cluster* yang optimal sudah tercapai (Pramana., dkk, 2018).

Gap statistik adalah teknik yang digunakan untuk menentukan jumlah *cluster* optimal dalam dataset. Metode ini membandingkan nilai variansi dalam *cluster* (*within-cluster variation*) antara dataset asli dengan distribusi acak data referensi. Kemudian, *clustering* dilakukan pada dataset acak dengan jumlah *cluster* yang sama seperti dataset asli. *Gap statistik* dihitung sebagai perbedaan antara nilai rata-rata log dari variansi dalam *cluster* pada dataset asli dengan nilai rata-rata log dari variansi dalam *cluster* pada dataset referensi acak. Jumlah *cluster* yang optimal dipilih ketika nilai *Gap statistik* mencapai puncaknya (Pramana., dkk, 2018).

Teknik *clustering* terdiri dari dua metode, yaitu *hirarki* dan *non hirarki*. Metode *hirarki* membentuk *cluster* secara *hirarkis* dengan menggabungkan *cluster* yang memiliki jarak terdekat atau memecah *cluster* yang lebih besar menjadi beberapa *cluster* yang lebih kecil. Sementara itu, metode *non hirarki* (*partitional*) tidak memiliki struktur *hirarki*. Beberapa Teknik *cluster* hirarki diantaranya adalah HDBSCAN, *single linkage*, *Aglomeratif*, *complete linkage*, *average linkage*, dan *Ward's method*. dan beberapa Teknik *cluster non hirarki* yang bisa digunakan adalah *K-Means*, *K-Medoids*, DBSCAN. Dalam penelitian ini, digunakan data *Covid-19* di Indonesia yang mengandung *outlier* karena beberapa daerah menunjukkan peningkatan kasus *covid-19* yang sangat tinggi dibandingkan daerah

lainnya seperti DKI Jakarta, Jawa barat, Jawa Tengah dan Jawa timur. Sehingga dalam penelitian ini digunakan teknik *clustering K-Medoids* yang merupakan salah satu metode *non hirarki* yang mengatasi adanya outlier (Suyanto, 2018). Dengan menggunakan *Medoid* sebagai pusat *cluster*, *K-Medoids* lebih tahan terhadap noise dan data pencilan (*outlier*) dibandingkan *K-Means*. Selain itu, *K-Medoids* juga dapat digunakan untuk data yang tidak berdistribusi normal atau tidak memiliki varian yang sama di semua dimensi. Namun, *K-Medoids* memerlukan lebih banyak waktu komputasi dibandingkan *K-Means* karena perhitungan jarak yang lebih kompleks.

K-Medoids adalah algoritma serupa dengan *K-Means* dalam pengelompokan data menjadi kelompok atau *cluster*. Perbedaannya terletak pada penentuan pusat *cluster* (Tibshirani, 2013). Dalam *K-Means*, pusat *cluster* ditentukan sebagai rata-rata aritmatika dari semua data dalam *cluster*. Sedangkan dalam *K-Medoids*, pusat *cluster* ditentukan sebagai salah satu data point yang paling representatif di dalam *cluster* tersebut.

Dalam *K-Medoids*, pemilihan *Medoid* di setiap iterasi dilakukan secara acak, kemudian dilakukan penghitungan jarak dari setiap data point ke *Medoid* yang terpilih. Setelah itu, data point yang memiliki total jarak *minimum* terhadap *Medoid* akan termasuk ke dalam *cluster* yang sesuai. Kemudian, *Medoid* baru dihitung untuk setiap *cluster* yang terbentuk, dan iterasi dilakukan kembali sampai tidak terdapat perubahan *cluster* yang signifikan.

Sebelumnya, telah dilakukan penelitian yang memfokuskan pada penerapan metode *K-Medoids* untuk mengelompokkan persebaran kasus *Covid-19* di Indonesia. Sebagai contoh, Gusmantoni (2020) melakukan penelitian tentang penerapan data mining dengan menggunakan *K-Medoids* untuk data *Covid-19* di Indonesia, dengan hasil bahwa terdapat tiga *cluster* kasus *Covid-19* di Indonesia. Hasil penelitian tersebut menjabarkan bahwa terdapat 32 provinsi pada *cluster* 1, satu provinsi pada *cluster* 2, dan dua provinsi pada *cluster* 3. Penelitian yang dilakukan oleh Ririn (2021) menunjukkan bahwa penerapan algoritma *K-Medoids* dalam data mining dapat membantu untuk mengelompokkan data kasus *Covid-19* di Jawa Barat dan memprediksi *cluster* wilayah yang terkena virus *Covid-19*. Hasil

penelitian, uji coba, dan implementasi menunjukkan bahwa data *Covid-19* di Jawa Barat dapat dikelompokkan menjadi tiga *cluster*. *Cluster* pertama (*cluster 0*) terdiri dari enam kabupaten/kota. *Cluster* kedua (*cluster 1*) terdiri dari enam belas kabupaten/kota dengan tingkat penyebaran *Covid-19* yang sedang. Sementara itu, *cluster* ketiga (*cluster 2*) memiliki tingkat penyebaran *Covid-19* yang rendah dan terdiri dari dua kabupaten/kota. Sedangkan penelitian menggunakan metode *K-Medoids* juga dilakukan oleh Samudi dkk., (2022) untuk pengelompokan data persebaran virus *corona* di DKI Jakarta. Berdasarkan hasil penelitian tersebut, ditemukan bahwa algoritma *K-Medoids* efektif dalam melakukan pengelompokan data *Covid-19*. Dalam analisis tersebut, terbentuk tiga kelompok atau *cluster*, dengan jumlah item pada *cluster 0* sebanyak 85, *cluster 1* sebanyak 123, dan *cluster 2* sebanyak 59. Total keseluruhan item yang dianalisis adalah 267 sesuai dengan data yang digunakan.

Penelitian-penelitian tersebut menunjukkan bahwa metode *K-Medoids* efektif dalam pengelompokan dan pemahaman persebaran virus *corona* di suatu daerah, namun jumlah *cluster* yang digunakan sudah ditentukan sebelumnya. Penggunaan metode *K-Medoids* pada data persebaran virus *Corona* membantu pemahaman dan pembuatan kebijakan yang tepat dalam penanganan virus.

Berdasarkan runtutan permasalahan diatas, maka penulis melakukan penelitian dengan judul **Penerapan Metode *K-Medoids* Menggunakan *Within cluster variation* Dan *Gap statistik* Berdasarkan Perkembangan Data *Covid-19* di Indonesia.**

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana mendapatkan jumlah *cluster* optimal untuk pengelompokan provinsi di Indonesia berdasarkan perkembangan data kasus *covid-19* di Indonesia dengan menggunakan *within cluster variation* dan *Gap statistik*?
2. Bagaimana Pengelompokan provinsi di Indonesia untuk data penyebaran data *covid-19* sebelum dan sesudah pemberian vaksin berdasarkan hasil *K-Medoids* dengan metode *within cluster variation* dan *gap statistik*?

3. Manakah metode terbaik untuk pengclusteran menggunakan analisis *K-Medoids* berdasarkan nilai *average silhouette width*?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data *covid-19* sebelum dan sesudah vaksin dengan variabel total kasus positif, total kasus meninggal, total kasus sembuh. Data yang diambil sebelum vaksin (26 Maret 2020 – 29 Februari 2021) dan data sesudah vaksin (1 Maret – 31 Desember 2022).
2. Sebelum melakukan analisis *K-Medoids* data terlebih dahulu di normalisasi menggunakan *min-max normalization*.
3. Hasil *cluster* di validasi menggunakan metode *average silhouette width*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian adalah sebagai berikut:

1. Mendapatkan *cluster* optimal untuk pengelompokan provinsi di Indonesia menggunakan *within cluster variation* dan Gap statistik.
2. Memperoleh Pengelompokan provinsi di Indonesia untuk data penyebaran data *covid-19* sebelum dan sesudah pemberian vaksin berdasarkan hasil *K-Medoids* dengan metode *within cluster variation* dan gap statistik.
3. Mendapatkan metode pengclusteran optimal terbaik berdasarkan nilai *average silhouette width*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah Penelitian ini diharapkan dapat memberikan masukan bagi pemerintah dan masyarakat dalam mengatasi serta memerangi virus *covid-19* ataupun wabah lainnya di masa yang akan datang, dan sebagai Literasi bagi masyarakat mengenai penyebaran *covid-19* di Indonesia.

BAB II

TINJAUAN PUSTAKA

2.1 *Clustering*

Penggunaan istilah "*cluster*", "*group*", dan "*class*" telah dilakukan secara intuitif tanpa adanya upaya untuk memberikan definisi formal. Bahkan, memberikan definisi formal bisa menjadi hal yang sulit dan mungkin tidak tepat. Bonner (1964) telah menyatakan bahwa kriteria utama dalam mengevaluasi arti dari istilah tersebut adalah mempertimbangkan nilai penggunaannya. Jika penggunaan istilah seperti "*cluster*" dapat memberikan hasil yang bernilai bagi peneliti, maka hanya itu yang dibutuhkan (Brian., dkk, 2011).

Clustering merupakan teknik dalam data mining yang bertujuan untuk mengelompokkan data-data (objek) ke dalam beberapa *cluster* atau kelompok sehingga objek dengan pola serupa disatukan kedalam *cluster* yang sama, sedangkan objek dengan pola berbeda harus menjadi bagian dari *cluster* yang berbeda. Adapun tujuan utama dari *clustering* adalah untuk memaksimalkan tingkat kesamaan pada sebuah *cluster* dan perbedaan pada *cluster* yang berbeda. Data yang dikelompokkan menggunakan metode ini tidak untuk diprediksi, melainkan hanya mengelompokkan data dengan struktur serupa (Nana & Mustakim, 2021).

Banyaknya metode pengelompokan diperlukan karena definisi "*cluster*" tidak memiliki definisi yang pasti, seperti yang dijelaskan oleh Estivill-Castro (2000). Karena itu, banyak metode pengelompokan telah dikembangkan dengan prinsip-prinsip induksi yang berbeda. Farley dan Raftery (1998) membagi metode pengelompokan menjadi dua kelompok utama, yaitu metode hirarkis dan partisi, sedangkan Han dan Kamber (2001) mengategorikan metode pengelompokan ke dalam tiga kategori tambahan, yaitu metode berbasis kepadatan, pengelompokan berbasis model, dan metode berbasis grid. Terdapat juga kategori-kategori alternatif yang membagi berbagai metode pengelompokan berdasarkan prinsip-prinsip induksinya, yang dapat ditemukan dalam karya Estivill-Castro (2000) (Rokarch, 2014).

Septianingsih (2022) didalam penelitiannya membagi *clustering* menjadi 2 metode yaitu metode *hirarki* dan metode *non hirarki*.

1. Metode Hirarki

Metode hirarki adalah metode *clustering* yang mana *cluster* dibentuk dengan secara hirarki dengan menggunakan konsep penggabungan dua *cluster* yang memiliki jarak terdekat atau pemecahan *cluster* yang lebih besar ke beberapa *cluster* yang lebih kecil. Pembentukan *cluster* pada metode *hirarki* pada umumnya menggunakan bagan atau *dendogram*.

2. Metode non Hierarki

Metode non hirarki (partitional) adalah metode pengelompokan ke dalam beberapa *cluster* tanpa adanya struktur hirarki. Pada metode ini setiap *cluster* memiliki titik pusat *cluster* yang disebut centroid. Proses *clustering* akan dilakukan secara berulang (iterasi) sehingga setiap objek (anggota) akan berpindah-pindah *cluster* hingga berakhir pada iterasi terakhir yaitu ketika semua anggota berada pada *cluster* yang memiliki jarak terdekat diantara semua *centroid* masing-masing *cluster*. Contoh metode non hirarki yang terkenal adalah *K-Means Clustering* *K-Medoids* dan CLARA.

2.1.1 *K-Medoids*

Algoritma *K-Medoids* menggunakan Teknik berbasis objek *representif* (perwakilan) yang disebut *Medoid* untuk mengatasi kelemahan *k-means* yang sensitive terhadap derau atau pencilan, dengan cara menghilangkan penggunaan rata-rata untuk memperbaharui *centroid* dan menggantinya dengan objek actual sebagai representasi dari suatu *cluster*. Jadi, algoritma *K-Medoids* melakukan partisi dengan cara meminimalkan jumlah *dissimilarity* antara setiap objek p dan objek representasi terdekat, yaitu menggunakan jumlah kesalahan *absolut* (Suyanto, 2018).

Konsekuensi dari strategi ini adalah kompleksitas yang menjadi lebih tinggi, karena harus melakukan pengecekan apakah penggantian objek *representatif* dengan setiap objek lain yang bukan *representatif* akan meningkatkan kualitas *clusterisasi*. Pada umumnya, *K-Medoids* diimplementasikan menggunakan

algoritma *Partitioning Around Medoid* (PAM), yang pada dasarnya mirip dengan algoritma *k-means* (Suyanto, 2018).

Langkah-langkah algoritma *K-Medoids* sebagai berikut (Munandar, 2022):

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*).
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan ukuran jarak *Euclidean distance* dengan persamaan:

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2} \quad (2.1)$$

Keterangan:

d_{ij} : nilai *Euclidean distance*

x_{ia} : variable ke a dari obyek i ($i = 1, \dots, n; a = 1, \dots, p$)

x_{ja} : variable ke a dari obyek j ($j = 1, \dots, n; a = 1, \dots, p$)

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *Medoid* baru.
4. Hitung jarak setiap objek yang berada pada setiap masing-masing *cluster* dengan menempuh *Medoid* baru.
5. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama.

Jika $S < 0$, maka ganti objek dengan data *cluster* untuk memperoleh sekelompok k objek yang baru sebagai *Medoid*. Ulangi tahap ke 3 sampai 5 hingga tidak terjadi perubahan *Medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing. Kemudian untuk mendapatkan nilai k di sebuah data yang ada di *clustering K-Medoids* dilakukan di dalam proses *clustering* dapat dipilih dengan bersasarkan nilai *within cluster variation* dan *Gap statistik*.

2.1.2 Jumlah *cluster* optimal

Jumlah optimal *cluster* (jumlah optimal *cluster*) merujuk pada jumlah *cluster* yang paling sesuai atau optimal untuk suatu data dalam analisis *clustering*.

Tujuan dalam menentukan jumlah optimal *cluster* adalah untuk memperoleh pengelompokan yang bermakna dan informatif dari data yang tersedia. Dalam penelitian ini, untuk menentukan jumlah optimasi *cluster*, digunakan metode evaluasi *within cluster variation dan Gap statistik*. Sebelum melakukan penentuan jumlah optimal *cluster*, data perlu dinormalisasi terlebih dahulu untuk memastikan keseragaman skala dan meminimalkan pengaruh variabel yang berbeda dalam analisis *clustering*.

2.1.2.1. Normalisasi Data

Menurut Suyanto (2018), teknik normalisasi data sangat penting dalam data mining, khususnya dalam proses klasifikasi dan *clusterisasi*. Normalisasi data dilakukan untuk mengatasi perbedaan rentang nilai atribut data, sehingga semua atribut memiliki bobot yang sama dan proses data mining tidak bias. Biasanya, normalisasi data dilakukan dengan mengubah rentang nilai atribut data menjadi rentang yang lebih kecil yaitu [0, 1]. Dalam penelitian ini, metode *K-Medoids* yang digunakan sudah dinormalisasi menggunakan *min-max normalization* dengan persamaan:

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\maxbaru_A - \minbaru_A) + \minbaru_A \quad (2.2)$$

Keterangan:

\min_A : Ini adalah nilai absolut *minimum* A.

\max_A : Ini adalah nilai absolut *maksimum* dari A.

\maxbaru_A : 1

\minbaru_A : 0

v' : Ini adalah nilai baru dari setiap data atribut.

v : Ini adalah nilai lama dari setiap data atribut.

Setelah data dinormalisasi, langkah selanjutnya adalah menentukan jumlah optimal *cluster* dengan menggunakan evaluasi *within cluster variation dan Gap statistik*.

2.1.2.2. *Within cluster variation*

Konsep *within cluster variation* menekankan perbedaan data pada setiap *cluster*. Tujuannya adalah untuk meminimalkan variasi data di setiap *cluster* sehingga setiap *cluster* memiliki data yang *homogen*. Dalam hal ini, data yang berada pada satu *cluster* harus memiliki karakteristik yang serupa dan berbeda dengan *cluster* lainnya. *Within cluster variation* dapat digunakan sebagai kriteria untuk menentukan jumlah *cluster* yang optimal dalam suatu dataset.

Penentuan jumlah *cluster* optimal sangat penting dalam penelitian menggunakan *K-Medoids*, karena jumlah *cluster* yang tidak sesuai akan menghasilkan *cluster* yang tidak *representatif* dan tidak memenuhi tujuan dari analisis. Oleh karena itu, penentuan jumlah *cluster* optimal sangat penting dan harus dilakukan dengan benar dan cermat.

Within Cluster Variaton didefenisikan sebagai berikut (Pramana., dkk, 2018):

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r \tag{2.3}$$

Dimana

$$D_r = \sum_{i \in C_r} \sum_{j \in C_r} \|x_i - x_j\|^2 \tag{2.4}$$

Keterangan:

- $\sum_{i \in C_r}$: Menunjukkan penjumlahan yang dilakukan untuk setiap titik data i yang termasuk dalam *cluster* r.
- $\sum_{j \in C_r}$: Menunjukkan penjumlahan yang dilakukan untuk setiap titik data j yang juga termasuk dalam *cluster* r.
- D_r : Merupakan variabel yang menunjukkan *within-cluster variation* untuk *cluster* r.
- $\frac{1}{2n_r}$: Merupakan faktor normalisasi yang digunakan dalam rumus. Faktor ini mengoreksi untuk jumlah titik data dalam *cluster* r, yang direpresentasikan oleh n_r .

$\|x_i - x_j\|^2$: Merupakan jarak kuadrat antara dua titik data i dan j dalam *cluster* r . Jarak ini dapat dihitung menggunakan metrik jarak tertentu, seperti *Euclidean distance*.

2.1.2.3. Gap statistik

Gap statistik membandingkan jumlah variasi dalam *cluster* dari data asli dengan data referensi *null distribution*. Nilai *within cluster variation* dari data referensi *null distribution* harus lebih besar dari nilai *within cluster variation* dari data asli untuk menentukan bahwa jumlah *cluster* tersebut optimal. *Gap statistik* membantu meminimalisasi bias dan *overfitting* yang sering terjadi pada metode *K-Medoids*, sehingga memberikan hasil yang lebih akurat dan objektif. Proses Algoritma *Gap statistik* sebagai berikut (pramana dkk., 2018):

1. Lakukan *monte carlo sampling* sebanyak B dari data
2. Untuk setiap nilai $1:(k)$,
 - a. Lakukan *clustering* dari data asli dan hitung nilai: $\log W_k$.
 - b. Lakukan *clustering* dari data hasil *monte carlo sampling* dan hitung nilai: $\log W_{kb}$.
 - c. Kemudian hitung nilai *Gap statistik*:

$$Gap_{(k)} = \frac{1}{B} \sum_{b=1}^B \log (W_{kb}) - \log (W_k) \quad (2.5)$$

Keterangan:

$Gap_{(k)}$: Gap untuk jumlah *cluster* k tertentu.

B : Jumlah bootstrap yang digunakan dalam metode *Gap statistik*.

W_{kb} : Dispersi intra-*cluster* untuk jumlah *cluster* k tertentu pada bootstrap ke- b .

W_k : Hasil *within cluster variation* untuk jumlah *cluster* k tertentu pada data asli.

2.1.3 Average Silhouette Width

Salah satu metode validasi pengujian hasil *clusterisasi* adalah *average Silhouette* yang diperkenalkan oleh Rousseeuw (1987). *Average Silhouette* adalah

suatu nilai yang mampu mengukur tingkat *kehomogenan* anggota dalam internal *cluster* dan tingkat *keheterogenan* anggota antar *cluster*. Nilai *Average Silhouette* berada pada interval -1 hingga 1. Semakin mendekati 1 nilai *Average Silhouette Width* maka menunjukkan bahwa hasil validasi *cluster* semakin baik dan sebaliknya jika mendekati -1 maka strukturnya buruk (Batool & Hennig, 2021).

Untuk menghitung nilai dapat menggunakan persamaan berikut:

$$ASW = \frac{1}{n} \sum_{i=1}^n S(i) \quad (2.6)$$

$$S(i) = \left(\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right) \quad (2.7)$$

$$a(i) = \frac{1}{n_c - 1} \sum_{r=1}^{n_c-1} d(x_{ic} - x_{rc}) \quad (2.8)$$

$$b(i) = \min d(i, C) \quad (2.9)$$

$$d(i, C) = \frac{1}{n_k} \sum_{r=1}^{n_k} d(x_i - x_{rc}) \quad (2.10)$$

Keterangan:

ASW : *Average Silhouette Width*

$S(i)$: Nilai *Silhouette* masing-masing objek

$a(i)$: jarak rata-rata sampel i ke sampel lain dalam *cluster*, dan

$b(i)$: jarak *minimum* sampel dari sampel i ke *cluster* lain.

n_c : jumlah data dalam *cluster* c

n_k : jumlah data di luar *cluster* c

$d(x_{ic} - x_{rc})$: jarak data ke i dengan data ke r dalam satu *cluster* j

$d(x_i - x_{rc})$: jarak data ke i dengan data ke r tidak satu *cluster* j

Tabel 2.1 Kriteria penilaian *clustering* berdasarkan *Average Silhouette Width*
(Guntara & Suprawoto, 2022)

Nilai average silhouette width	Kriteria
0,71 - 1,00	Struktur Kuat
0,51 - 0,70	Struktur baik
0,25 - 0,50	Struktur Lemah
$\leq 0,25$	Struktur buruk

2.2 Covid-19

Pertengahan Desember 2019, Virus misterius mulai berkembang di pasar ikan daerah Wuhan, Provinsi Hubei, China. Dilaporkan bahwasanya Sejak 31 Desember 2019 hingga 3 Januari 2020 kasus ini meningkat pesat, ditandai dengan dilaporkannya sebanyak 44 kasus. Badan Kesehatan Dunia yaitu WHO mengumumkan, bahwa perkembangan kasus virus di China akan menimbulkan virus varian baru yaitu coronavirus jenis baru (SARS-CoV-2). Awalnya, penyakit ini dinamakan sementara sebagai *2019 novel coronavirus* (2019-nCoV), kemudian WHO mengumumkan nama baru pada 11 Februari 2020 yaitu *Coronavirus Disease (COVID-19)* yang disebabkan oleh virus *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-CoV-2) (Nana & Mustakim, 2021).

Banyak coronavirus diketahui beredar dan ada pada hewan yang belum terinfeksi oleh manusia. Tanda-tanda khas infeksi termasuk gangguan pernapasan, demam, batuk, sesak napas, dan kesulitan bernapas. China melaporkan di Wuhan yang kemudian menyerang negara-negara di hampir seluruh dunia, termasuk Asia Tenggara. *Covid-19* umumnya merupakan penyakit yang dapat disembuhkan tetapi juga dapat mematikan. Berdasarkan data WHO pada 30 April 2020, ada sekitar 3.090.445 kasus yang dikonfirmasi dan lebih dari 217.769 orang meninggal akibat virus *Covid-19* (Hutagulung., dkk, 2021).

Berdasarkan data dari situs resmi kementerian kesehatan Republik Indonesia (*kemkes.go.id*), situs pemerintah tentang *Covid-19* (*covid19.go.id*), dan Badan Nasional Penanggulangan Bencana (BNPB), telah terdeteksi bahwa virus

Covid-19 sudah masuk di Indonesia pada tanggal 2 Maret 2020 di Provinsi Jawa barat. Hingga Sabtu (31/12/2022), jumlah orang yang terkonfirmasi positif virus corona di Indonesia telah mencapai 6.719.815, serta 160.612 orang meninggal, 9.871 orang positif aktif (dirawat), dan 6.549.332 orang dinyatakan sembuh.

Dalam hal ini, Penegakan Hukum di awal munculnya virus di Indonesia. Pemerintah Indonesia di tinjau berdasarkan Pasal 154 Undang – Undang Nomor 36 tahun 2009 tentang Kesehatan, menyatakan Pemerintah wajib mengumumkan bagian wilayah yang menjadi sumber terjangkitnya penularan penyakit ke banyak masyarakat. Pemerintah wajib mengungkap jenis dari penyakit yang penularannya menyebar dengan cepat (Sukur., dkk, 2020).

Fakta dari Pemerintah dalam melindungi jaminan kesehatan masyarakat dikatakan lamban untuk menyebarkan informasi terkait kasus yang memakan korban banyak karena adanya virus yang sangat berbahaya ini. Sehingga dalam Pelayanan Kesehatan yang dilakukan tenaga medis bisa dikatakan hampir tidak mampu karena disebabkan banyaknya pasien yang dinyatakan Positif *Covid-19*. Berdasarkan uraian diatas, maka akan dibahas mengenai bagaimana awal terjadinya penyebaran virus begitu cepat di inodnesia dan bagaimana pelayanan kesehatan untuk masyarakat Indonesia dengan adanya keterlambatan informasi dalam mengungkapkan terjadinya penyebaran virus *Covid-19* (Sukur., dkk, 2020).

Dalam hal mencegah penyebaran virus *Covid-19* alat yang paling ampuh untuk melakukan hal tersebut ialah Vaksin. Tujuan dari Vaksin ialah untuk memberikan kekebalan terhadap penyakit, sehingga apabila tidak terdapat manfaat dari vaksin tersebut maka tidak akan ada kekebalan terhadap tubuh pengguna vaksin dan lebih mudah terserang penyakit. Pengaturan terhadap Vaksin *covid19* terdapat dalam Peraturan Presiden No.99 Tahun 2020 tentang Pengadaan Vaksin dan Pelaksanaan Vaksinasi Dalam Rangka Penanggulangan Pandemi Corona Virus Disease 2019 (*Covid-19*) (Perpres Vaksin) dan Peraturan Menteri Kesehatan (PMK) Nomor 84 Tahun 2020 tentang Pelaksanaan Vaksinasi dalam Rangka Penanggulangan Pandemi *Covid-19* (PMK Vaksin). Perpres Vaksin sebagian besar mengatur mengenai kerjasama terhadap pendistribusian vaksin dan juga beberapa peran didalam melaksanakan pendistribusian vaksin di Indonesia (Sukawati, 2021).

Teknis dan pedoman terhadap penyebaran *Covid-19* diatur lebih lanjut dalam PMK Vaksin, pengaturan tersebut termasuk sasaran dari vaksin, jenis vaksin, jumlah vaksin, dll. Pengaturan terhadap vaksin tersebut mengindikasikan bahwa Vaksin *Covid-19* yang akan didistribusikan berasal dari WHO. Sehingga, dapat dikatakan Indonesia menunggu kedatangan vaksin dari organisasi dunia tersebut untuk didistribusikan. Akan tetapi, Indonesia dapat membuat vaksin tersendiri untuk diedarkan asalkan mendapat persetujuan dari WHO (Sukawati, 2021).

Vaksin ditujukan kepada setiap Warga Negara Indonesia baik yang berada diluar negeri maupun dalam negeri, dikarenakan setiap Warga Negara Indonesia memiliki hak yang sama terutama dalam perlindungan penyakit yang mematikan. Pengedaran vaksin kepada masyarakat dilakukan secara bertahap, bahkan terdapat golongan prioritas untuk penerima Vaksin *Covid-19*. Sebagaimana yang dijelaskan didalam pasal 8 ayat (4) PMK Vaksin bahwa yang termasuk didalam golongan prioritas yaitu tenaga kesehatan, TNI, Kepolisian Negara Republik Indonesia, aparat hukum, tokoh masyarakat, pelaku perekonomian, perangkat daerah, tenaga pendidik, aparatur kementerian/lembaga, dan masih banyak yang lainnya. Jika dilihat pada penjelasan tersebut bahwasannya golongan prioritas wajib diberikan vaksin paling pertama termasuk golongan prioritas yang berada di luar negeri (Sukawati, 2021).