

**IMPLEMENTASI ALGORITMA K-MEANS DALAM
MENENTUKAN TINGKAT PENYEBARAN
COVID-19 DI INDONESIA**

SKRIPSI



disusun oleh
Elsa Virantika
18.11.2067

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**IMPLEMENTASI ALGORITMA K-MEANS DALAM
MENENTUKAN TINGKAT PENYEBARAN
COVID-19 DI INDONESIA**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Elsa Virantika
18.11.2067

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA K-MEANS DALAM
MENENTUKAN TINGKAT PENYEBARAN
COVID-19 DI INDONESIA

yang dipersiapkan dan disusun oleh

Elsa Virantika

18.11.2067

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Juni 2022

Dosen Pembimbing,

Kusnawi, S.Kom., M.Eng.
NIK. 190302112

PENGESAHAN
SKRIPSI
IMPLEMENTASI ALGORITMA K-MEANS DALAM
MENENTUKAN TINGKAT PENYEBARAN
COVID-19 DI INDONESIA

yang dipersiapkan dan disusun oleh

Elsa Virantika

18.11.2067

telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Juni 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Supratni, M.Kom
NIK. 190302239

Majid Rahardi, S.Kom., M.Eng
NIK. 190302393

Kusnawi, S.Kom, M.Eng.
NIK. 190302112

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer.
Tanggal 27 Juni 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 30 Juni 2022



Elsa Virantika

NIM. 18.11.2067

MOTTO

JIKA ALLAH MENOLONG KAMU, MAKA TIDAK ADA YANG DAPAT MENGALAHKANMU, TETAPI JIKA ALLAH MEMBLARKAN KAMU (TIDAK MEMBERI PERTOLONGAN), MAKA SIAPA YANG DAPAT MENOLONGMU SETELAH ITU? KARENA ITU, HENDAKLAH KEPADA ALLAH SALA ORANG-ORANG MUKMIN BERTAWAKAL.

[QS. ALI IMRAN:160]

BOLEH JADI KAMU MEMBENCI SESUATU, PADAHAL IA AMAT BAIK BAGIMU, DAN BOLEH JADI (PULA) KAMU MENYUKAI SESUATU, PADAHAL IA AMAT BURUK BAGIMU. ALLAH MENGETAHUI, SEDANG KAMU TIDAK MENGETAHUI.

[QS. AL BAQARAH 216]

SEBENARNYA ALLAH TIDAK AKAN MENGUBAH KEADAAN SUATU KAUM, KEUALAI MEREKA MENGUBAH KEADAAN MEREKA SENDIRI.

[QS. AR RA'D 11]

ALLAH TIDAK MEMBEANI SESEORANG MELAINKAN SESUAI KESANGGUPANNYA.

[QS. AL BAQARAH 286]

PERSEMBAHAN

Alhamdulillah Puji Syukur ini penulis panjatkan, akhirnya skripsi ini dapat terselesaikan dengan baik. Karya ini merupakan wujud dari kegigihan dalam ikhtiar untuk sebuah makna kesempurnaan dengan tanpa ebrharap melampaui kemaha sempurna sang maha sempurna. Selaku penulis mempersembahkan skripsi ini kepada:

1. Allah SWT atas ridho-Nya, skripsi ini dapat terselesaikan dengan baik, Sujud syukur saya panjatkan kepada-Mu dan jadikanlah hamba-Mu ini seseorang yang pandai bersyukur dan selalu dalam lindungan-Mu.
2. Shalawat serta salam senantiasa tercurah kepada Nabi Agung Muhammad SAW, Sebagai sang pencerah yang menyempurnakan akhlak manusia menjadi manusia yang lebih cerdas.
3. Tidak bisa dipungkiri kedua orang tua saya tercinta, yaitu Slamet Widodo dan Neni Indriyasaki menjadi salah satu motivasi terbesar untuk menyelesaikan skripsi ini. Yang selalu memanjatkan doa untuk putrinya dalam setiap sujudnya, yang selalu menasehati, menjadi mood booster dikala putri kecilmu ini hampir menyerah. Terima Kasih Ibu... Terima Kasih Ayah atas semua yang telah engkau berikan kepada putri kecilmu ini, semoga selalu diberikan kesehatan dan panjang umur agar dapat menemani langkah kecilmu bersama adikku tercinta Arysa Ayla Virantika menuju kesuksesan.
4. Keluarga besarku, kakak-kakakku, keponakanku, terima kasih telah menjadi penyemangat untuk menjadikan ku sosok yang kuat dan tabah dalam menjalani kehidupan ini.

5. Untuk orang paling istimewa dalam hidupku. Kamu adalah sosok terbaik, sosok penyabar, yang selalu memberiku motivasi, saran, dan dukungan tanpa henti yang tidak bisa tetap acuh pada masalah orang-orang yang membutuhkan bantuan. Terima kasih telah menjadi mood boosterku setelah kedua orang tuaku, betapa beruntungnya aku bertemu denganmu di jalan hidupku.
6. Teruntuk Bapak Kusnawi, S.Kom, M.Eng selaku dosen pembimbing, saya ucapkan terima kasih karena telah menjadi sosok guru yang teramat baik yang bersedia membimbing, yang sabar menuntun setiap langkah saya, selalu memberikan kritik dan saran yang membangun serta mengantarkan saya untuk mengantungi gelar sarjana.
7. Teman-teman 18 IF 04, terima kasih banyak untuk bantuan dan kerja samanya selama ini, teman seperjuangan saya selama ini.
8. Serta semua pihak yang tidak bisa saya sebutkan satu persatu yang telah mendoakan, mendukung dan memotivasi saya selama ini.

KATA PENGANTAR

Assalamu'alaikum Warrahmatullahi Wabarakatuh

Alhamdulillah rabbil'alamin, puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, hidayah, dan inayah-Nya, sehingga penulis dapat menyelesaikan laporan skripsi ini dengan judul **"Implementasi Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia"**, sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Jurusan Informatika Universitas Amikom Yogyakarta.

Dalam proses penyusunan laporan skripsi ini, penulis menyadari bahwa laporan skripsi ini tidak mungkin terselesaikan tanpa adanya bantuan, dukungan, bimbingan, dan nasehat dari berbagai pihak baik secara langsung maupun tidak langsung yang telah memberikan motivasi kepada penulis. Pada kesempatan ini, penulis menyampaikan terima kasih setulus-tulusnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta selaku Dekan Fakultas Ilmu Komputer
3. Ibu Windha Mega Pradnya Dhuhita, M.Kom selaku Ketua Program Studi Informatika di Universitas Amikom Yogyakarta.
4. Bapak Kusnawi, S.Kom, M.Eng, sebagai dosen pembimbing yang telah sangat banyak membantu saya dalam memberikan petunjuk, saran, pengarahan, serta bimbingan yang besar dalam skripsi ini.
5. Kedua orang tua penulis, Slamet Widodo dan Neni Indriyasari yang selalu memberikan kasih sayang, doa, nasehat, serta kesabarannya yang luar biasa

dalam setiap langkah hidup penulis, yang banyak memberikan dukungan moril serta materil dengan tulus ikhlas.

6. Adik penulis tercinta, Arsyia Ayla Virantika, terima kasih atas doa dan segala dukungan.
7. Teman-teman yang telah ikut andil dan banyak membantu dalam penyelesaian skripsi ini.
8. Seluruh staff dan karyawan Universitas Amikom Yogyakarta yang banyak membantu kelancaran segala aktivitas penulisan skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan kritik dan saran yang membangun selalu penulis harapkan demi kemajuan dan menuju arah yang lebih baik di masa yang akan datang sehingga dapat bermanfaat bagi penulis serta pihak-pihak yang membutuhkan. Semoga laporan skripsi ini bermanfaat bagi semua pihak.

Wassalamu 'alaikum Warrahmatullah Wabarakatuh

Yogyakarta, 27 Juni 2022

Elsa Virantika

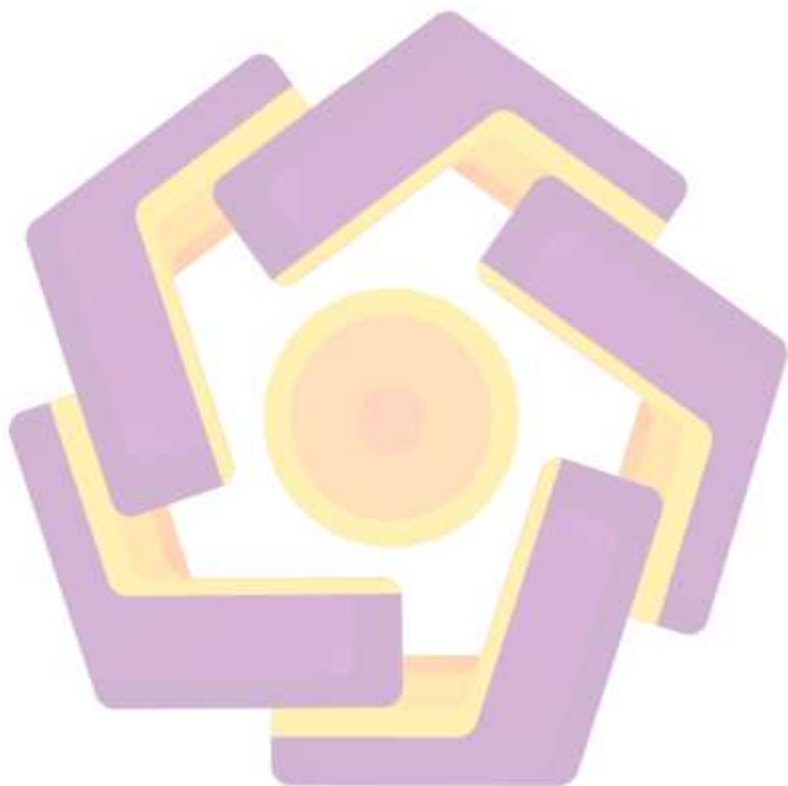
18.11.2067

DAFTAR ISI

JUDUL.....	I
PERSETUJUAN.....	II
PERNYATAAN.....	III
MOTTO.....	V
PERSEMBAHAN.....	VI
KATA PENGANTAR.....	VIII
DAFTAR ISI.....	X
DAFTAR TABEL.....	XIV
DAFTAR GAMBAR.....	XVI
INTISARI.....	XVII
ABSTRACT.....	XVIII
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH.....	3
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	4
1.5 MANFAAT PENELITIAN.....	5
1.6 METODE PENELITIAN.....	6
1.6.1 Identifikasi Masalah.....	6
1.6.2 Metode Pengumpulan Data.....	6
1.6.3 Metode Analisis.....	6
1.7 SISTEMATIKA PENULISAN.....	7
BAB II LANDASAN TEORI.....	9
2.1 TINJAUAN PUSTAKA.....	9
2.1.1 Penelitian Terdahulu.....	11
2.1.2 Deskripsi Jurnal.....	14

2.2	<i>DATA MINING</i>	14
2.2.1	<i>Teknik Pembelajaran Data Mining</i>	17
2.2.2	<i>Fungsi Data Mining</i>	18
2.3	<i>CLUSTERING</i>	19
2.3.1	<i>Metode Clustering Data Mining</i>	20
2.3.1.1	<i>Metode Partitioning Clustering</i>	20
2.3.1.2	<i>Metode Hierarchical Clustering</i>	21
2.3.1.3	<i>Metode Grid-Based Clustering</i>	22
2.3.1.4	<i>Metode Density-Based Clustering</i>	23
2.3.1.5	<i>Metode Distribution Based Clustering</i>	24
2.3.1.6	<i>Metode Fuzzy Clustering</i>	24
2.3.1.7	<i>Metode Constraint-Based Clustering</i>	25
2.4	<i>ALGORITMA K-MEANS</i>	25
2.5	<i>DATA SCALING</i>	29
2.5.1	<i>Robust Scaler</i>	29
2.6	<i>PENGUJIAN TINGKAT CLUSTERISASI</i>	30
2.6.1	<i>Elbow Method</i>	30
2.6.2	<i>DB Index</i>	30
2.6.3	<i>Silhouette Coefficient</i>	32
2.7	<i>COVID-19</i>	33
2.7.1	<i>Probable</i>	34
2.7.2	<i>Konfirmasi atau positif</i>	35
2.7.3	<i>Kematian</i>	35
2.8	<i>PYTHON</i>	35
BAB III METODOLOGI PENELITIAN		36
3.1	ALAT DAN BAHAN	36
3.1.1	<i>Perangkat Keras</i>	36
3.1.2	<i>Perangkat Lunak</i>	36
3.2	TAHAPAN PENELITIAN	37
3.2.1	<i>Identifikasi Masalah</i>	38
3.2.2	<i>Studi Literatur</i>	39
3.2.3	<i>Pengumpulan Data</i>	39

3.2.4	<i>Selection Data</i>	39
3.2.5	<i>Data Cleaning</i>	40
3.2.6	<i>Data Reduction</i>	43
3.2.7	<i>Data Transformation</i>	43
3.2.7.1	Kasus Harian	44
3.2.7.2	Kasus Sembuh	47
3.2.7.3	Kasus Meninggal Dunia Harian	50
3.2.8	<i>Data Mining</i>	55
3.2.8.1	<i>Clustering K-Means</i>	55
3.2.9	<i>Evaluation</i>	71
3.2.9.1	<i>Elbow Method</i>	71
3.2.9.2	<i>Davies Bouldin Index</i>	75
3.2.9.3	<i>Silhouette Coefficient</i>	77
BAB IV HASIL DAN PEMBAHASAN		81
4.1	IMPLEMENTASI CLUSTERING K-MEANS	81
4.2	PENGUJIAN	100
4.3	PROSEDUR PENGUJIAN	101
4.3.1	<i>Prosedur Pengujian Algoritma K-means</i>	101
4.4	DATA HASIL PENGUJIAN	101
4.4.1	<i>Hasil Pengujian Algoritma K-means</i>	101
4.4.1.1	<i>Elbow method</i>	102
4.4.1.2	<i>Davies Bouldin Index (DBI)</i>	105
4.4.1.3	<i>Silhouette Coefficient</i>	107
4.5	ANALISA DATA / EVALUASI	109
4.5.1	<i>Evaluasi Algoritma K-means</i>	109
4.5.2	<i>Evaluasi Hasil Pengujian Pemanding</i>	110
BAB V KESIMPULAN DAN SARAN		112
5.1	KESIMPULAN	112
5.2	SARAN	113
DAFTAR PUSTAKA		114



DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	11
Tabel 2. 2 Kriteria pengukuran <i>Silhouette Coefficient</i>	33
Tabel 3. 1 Data Awal.....	40
Tabel 3. 2 Perhitungan Rata-rata dan Standar Deviasi.....	42
Tabel 3. 3 Perhitungan <i>Threshold</i>	42
Tabel 3. 4 Hasil Transformasi Kasus Harian.....	45
Tabel 3. 5 Hasil Transformasi Kasus Sembuh.....	48
Tabel 3. 6 Hasil Transformasi Kasus Meninggal Dunia	51
Tabel 3. 7 Hasil Transformasi Kasus Harian, Kasus Sembuh, dan Kasus Meninggal Dunia	53
Tabel 3. 8 Titik Pusat Awal.....	56
Tabel 3. 9 Hasil Perhitungan <i>Euclidean Distance</i>	56
Tabel 3. 10 Titik Pusat Iterasi 2.....	58
Tabel 3. 11 Hasil Perhitungan Iterasi 2	59
Tabel 3. 12 Titik Pusat Iterasi 3.....	60
Tabel 3. 13 Hasil Perhitungan Iterasi 3	61
Tabel 3. 14 Titik Pusat Iterasi 4.....	62
Tabel 3. 15 Hasil Perhitungan Iterasi 4	63
Tabel 3. 16 Titik Pusat Iterasi 5.....	64
Tabel 3. 17 Hasil Perhitungan Iterasi 5	65
Tabel 3. 18 Titik Pusat Iterasi 6.....	66
Tabel 3. 19 Hasil Perhitungan Iterasi 6	67
Tabel 3. 20 Titik Pusat Iterasi 7.....	68
Tabel 3. 21 Hasil Perhitungan Iterasi 7	69
Tabel 3. 22 Perhitungan <i>Elbow Method</i>	72
Tabel 3. 23 <i>Squared Error</i> dan <i>Mean Variance</i>	74
Tabel 3. 24 Sum of Square Error.....	74
Tabel 3. 25 Hasil <i>Sum of square within cluster</i>	75
Tabel 3. 26 Hasil <i>Sum of square between cluster</i>	76
Tabel 3. 27 Hasil Pengukuran Rasio	76

Tabel 3. 28 Hasil <i>DB Index</i>	77
Tabel 3. 29 Perhitungan <i>Silhouette Coefficient</i>	78
Tabel 3. 30 Hasil Rata-rata <i>Silhouette Coefficient</i>	79
Tabel 3. 31 Hasil <i>Silhouette Coefficient</i>	80
Tabel 4. 1 Kode Program Pengambilan Data	81
Tabel 4. 2 Kode Program Visualisasi Atribut	82
Tabel 4. 3 Kode Program Data <i>Cleaning</i>	83
Tabel 4. 4 Kode Program <i>Feature Selection</i>	83
Tabel 4. 5 Kode Program Data <i>Reduction</i>	85
Tabel 4. 6 Kode Program Data Transformation	85
Tabel 4. 7 Kode Program Transformasi Kasus Harian	86
Tabel 4. 8 Kode Program Transformasi Kasus Sembuh	88
Tabel 4. 9 Kode Program Transformasi Kasus Meninggal Dunia	90
Tabel 4. 10 Hasil Transformasi Setiap Atribut	92
Tabel 4. 11 Kode Program <i>Clustering K-Means</i>	94
Tabel 4. 12 Titik Pusat Awal	95
Tabel 4. 13 Hasil <i>Clustering K-Means</i>	96
Tabel 4. 14 Clusterisasi tiap Provinsi	98
Tabel 4. 15 Kode Program Elbow Method	103
Tabel 4. 16 Hasil Elbow Method (Sum of Square Error)	104
Tabel 4. 17 Kode Program DB Index	105
Tabel 4. 18 Hasil <i>DB Index</i>	106
Tabel 4. 19 Kode Program <i>Silhouette Coefficient</i>	107
Tabel 4. 20 Hasil <i>Silhouette Coefficient</i>	109
Tabel 4. 21 Hasil Pengujian Pembandingan	110

DAFTAR GAMBAR

Gambar 2. 1 Proses <i>Knowledge Discovery in Database (KDD)</i> [16]	15
Gambar 2. 2 Fungsi <i>Data mining</i> [17].....	18
Gambar 2. 3 Ilustrasi Metode <i>Partitioning Clustering</i> [21].....	21
Gambar 2. 4 Ilustrasi Metode <i>Hierarchical Clustering</i> [21].....	22
Gambar 2. 5 Ilustrasi Metode <i>Density-Based Clustering</i> [21]	23
Gambar 2. 6 Ilustrasi Metode <i>Distribution-Based Clustering</i> [21].....	24
Gambar 2. 7 Ilustrasi Metode <i>Constraint-Based Clustering</i> [21]	25
Gambar 2. 8 Diagram Alur Algoritma <i>Clustering K-means</i> [24].....	27
Gambar 3. 1 Tahapan Penelitian.....	38
Gambar 3. 2 Histogram Kasus Harian Sebelum Transformasi	44
Gambar 3. 3 Histogram Kasus Harian Setelah Transformasi	45
Gambar 3. 4 Histogram Kasus Sembuh Sebelum Transformasi	47
Gambar 3. 5 Histogram Kasus Sembuh Sesudah Transformasi	47
Gambar 3. 6 Histogram Kasus Meninggal Sebelum Transformasi	50
Gambar 3. 7 Histogram Kasus Meninggal Dunia Setelah Transformasi	51
Gambar 4. 1 Grafik <i>Heatmap Person Correlation Matrix</i>	84
Gambar 4. 2 <i>Boxplot</i> Kasus Harian	88
Gambar 4. 3 <i>Boxplot</i> Kasus Sembuh.....	90
Gambar 4. 4 <i>Boxplot</i> Kasus Meninggal Dunia	92
Gambar 4. 5 Visualisasi <i>Subplot Clustering K-Means</i>	99
Gambar 4. 6 Peta Persebaran Kasus <i>Covid-19</i> di Indonesia.....	100
Gambar 4. 7 Hasil Uji <i>Elbow Method</i>	104
Gambar 4. 8 Hasil Uji <i>DB Index</i>	106
Gambar 4. 9 Hasil Uji <i>Silhouette Coefficient</i>	108

INTISARI

Coronavirus Diseases 2019, sering dikenal dengan *Covid-19*, adalah penyakit menular yang disebabkan oleh virus *SARS-CoV-2*. Indonesia memiliki wilayah yang luas sehingga mudah terjangkit *covid-19* serta persebaran virus *covid-19* di Indonesia berkembang cukup pesat. Berdasarkan wilayah di Indonesia dapat dikelompokkan ke dalam bagian-bagian provinsi di Indonesia serta menghasilkan titik-titik provinsi persebaran kasus *covid-19*, bertujuan untuk membuat strategi penanganan penyebaran *covid-19* di seluruh provinsi di Indonesia.

Pengelompokan tingkat penyebaran *covid-19* dilakukan menggunakan metode *data mining*, yaitu *algoritma k-means clustering*, dengan mengelompokkan data ke beberapa *cluster* berdasarkan kemiripan datanya. Berdasarkan hasil penelitian, teridentifikasi 3 *cluster*, yaitu *cluster 0* dengan tingkat persebaran *covid-19* kategori rendah terdapat 12 provinsi, *cluster 1* dengan tingkat persebaran *covid-19* kategori sedang terdapat 18 provinsi, dan *cluster 2* dengan tingkat persebaran *covid-19* kategori tinggi terdapat 4 provinsi.

Berdasarkan hasil penelitian tersebut, diharapkan dapat memberikan informasi serta mendukung pemerintah untuk mengambil keputusan secara strategis pada setiap *cluster* guna menekan tingkat persebaran *covid-19* di Indonesia.

Kata Kunci: *Data Mining*, *Algoritma K-Means*, Penyebaran, Pengelompokan

ABSTRACT

Coronavirus Diseases 2019, often known as Covid-19, is an infectious disease caused by the SARS-CoV-2 virus. Indonesia has a large area so that it is easy to contract COVID-19 and the spread of the Covid-19 virus in Indonesia is growing quite rapidly. Based on the region in Indonesia, it can be grouped into parts of provinces in Indonesia and produce provincial points for the distribution of Covid-19 cases, aiming to create a strategy for handling the spread of COVID-19 in all provinces in Indonesia.

The grouping of the level of spread of COVID-19 is carried out using a data mining method, namely the k-means clustering algorithm, by grouping data into several clusters based on the similarity of the data. Based on the results of the study, 3 clusters were identified, namely cluster 0 with a low level of distribution of Covid-19, 12 provinces, cluster 1 with a moderate level of distribution of COVID-19, 18 provinces, and cluster 2 with a high level of distribution of COVID-19, 4 categories. province.

Based on the results of this study, it is hoped that it can provide information and support the government to make strategic decisions in each cluster to reduce the level of spread of COVID-19 in Indonesia.

Keyword: *Data Mining, K-Means Algorithm, Spreading, Clustering*

