

# Klasifikasi Daerah Rawan Covid-19 Dengan Menggunakan Metode K Means Clustering Di Indonesia

Maria N.S Napitupulu<sup>1</sup>, Doli Hasibuan<sup>2</sup>, Marzuki Sinambela<sup>3</sup>

<sup>1,2,3</sup>Fakultas Ilmu Komputer, Universitas Methodist Indonesia

## Info Artikel

### Histori Artikel:

Received, Okt 19, 2021

Revised, Nov 01, 2021

Accepted, Des 19, 2021

### Keywords:

Covid-19

Data mining

K-means

## ABSTRAK

Wabah virus corona telah meluas hampir seluruh negara di dunia terutama di Indonesia. Masyarakat dianjurkan untuk tetap di rumah jangan keluar rumah kecuali dalam keadaan darurat. Di Indonesia mempunyai jumlah kasus terkonfirmasi, pulih, dan kematian ini sangatlah bervariasi dari tempat ke tempat, yang dikarenakan perbedaan kebijakan dan kemampuan pemerintah tiap-tiap tempat untuk melakukan tindakan pencegahan dan penindakan terhadap pandemi ini seperti social distancing, tracing dan testing. Namun karena ada banyaknya data yang tersedia, maka Big Data seperti ini sangatlah sulit untuk dimanfaatkan secara baik, maka digunakan data mining dengan menggunakan algoritma K-Means Clustering. Salah satu teknik yang dapat digunakan adalah teknik Clustering, terutamanya K-Means Clustering dan mengolah data COVID-19 yang terdapat pada situs resmi covid-19 di Indonesia yaitu covid19.go.id. Dengan menggunakan metode Data Mining yaitu K-Means Clustering, dimana data dikategorikan berdasarkan mean terdekat, sehingga daerah-daerah dapat dikategorikan menjadi kluster-kluster berdasarkan dampak COVID pada daerah tersebut, dan prioritas bantuan COVID dapat ditentukan dan diarahkan berdasarkan informasi kluster hasil pengelompokan, dimana zona merah terdiri dari 4 provinsi, zona kuning terdiri dari 11 provinsi, zona hijau terdiri dari 19 provinsi.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Penulis Koresponden:

Maria N.S Napitupulu, Faculty of  
Computer Science,  
Universitas Methodist Indonesia, Medan,  
Jl. Hang Tuah No.8, Medan - Sumatera Utara.  
Email: [napitupulumaria9@gmail.com](mailto:napitupulumaria9@gmail.com)

## 1. PENDAHULUAN

Wabah virus corona baru SARSCoV-2 (penyakit coronavirus 2019; sebelumnya 2019nCoV) yang mulai berkembang di Provinsi Hubei, Republik Rakyat Cina, telah menyebar ke banyak negara lain. Pada 30 Januari 2020 Komite Darurat WHO mengumumkan keadaan darurat kesehatan global berdasarkan tingkat pemberitahuan kasus yang meningkat di Cina dan lokasi internasional. Hingga pertengahan Februari 2020 Cina menanggung beban morbiditas dan mortalitas yang cukup besar, sedangkan di negara-negara Asia lainnya, di Eropa dan Amerika Utara sejauh ini masih rendah. Coronavirus adalah virus RNA besar berantai tunggal positif yang menginfeksi manusia, tetapi juga berbagai jenis hewan. Namun, upaya untuk mengidentifikasi inang perantara yang potensial tampaknya telah diabaikan di Wuhan dan rute penularan yang tepat sangat perlu diperjelas (Velavan & Meyer,

2020). Data jumlah kasus terkonfirmasi, pulih, dan kematian ini sangatlah bervariasi dari tempat ke tempat, yang dikarenakan perbedaan kebijakan dan kemampuan pemerintah tiap-tiap tempat untuk melakukan tindakan pencegahan dan penindakan terhadap pandemi ini seperti social distancing, tracing dan testing. Data-data masing-masing daerah ini dapat ditemukan dan diakses pada website resmi Covid-19 Indonesia yaitu Covid19.go.id. Namun karena ada banyaknya data yang tersedia, maka Big Data seperti ini sangatlah sulit untuk dimanfaatkan secara baik. Maka, agar dapat diperoleh informasi yang berguna dan mudah untuk digunakan, dapat digunakan teknik Data Mining untuk mengolah data dengan jumlah sebesar ini.

Salah satu teknik yang dapat digunakan adalah teknik Clustering, terutamanya K-Means Clustering, dimana data dikategorikan berdasarkan mean terdekat, sehingga daerah-daerah dapat dikategorikan menjadi kluster-kluster berdasarkan dampak COVID pada daerah tersebut, dan prioritas bantuan COVID dapat ditentukan dan diarahkan berdasarkan informasi kluster tersebut, sehingga penindakan terhadap pandemi COVID-19 ini dilakukan secara seefektif mungkin (Indraputra & Fitriana, 2017). Clustering adalah proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam subset, sehingga data dalam setiap subset memiliki arti yang bermanfaat (Merliana, Ernawati, & Santoso, 2018). Makapenulis membuat suatu penelitian untuk memberikan solusi alternatif yaitu dengan judul “Klasifikasi Daerah Rawan Covid-19 Dengan Menggunakan Metode K Means Clustering Di Indonesia”.

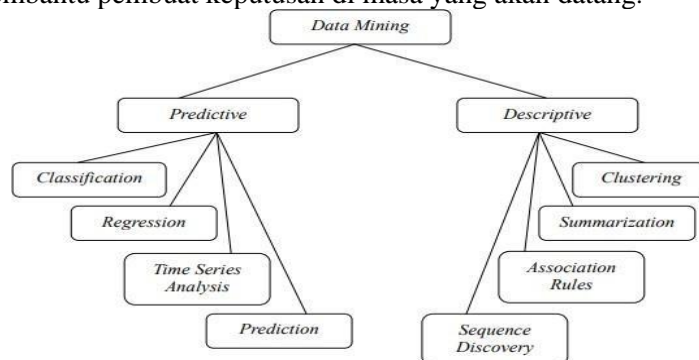
## 2. METODOLOGI PENELITIAN

### 2.1. Metode Pengumpulan Data

1. Studi Pustaka Pengumpulan data dilakukan dengan cara mempelajari, meneliti, dan menelaah berbagai literatur dari perpustakaan yang bersumber dari buku-buku, jurnal ilmiah, situs internet, dan bacaan lainnya yang berkaitan dengan penelitian yang dilakukan.
2. Pengamatan (Observasi) merupakan cara mengamati obyek penelitian untuk mengerti tentang kebutuhan obyek penelitian tersebut sehingga aplikasi yang dibangun dapat memenuhi kebutuhan *user* yang bersangkutan.

### 2.2. Data Mining

*Data mining* adalah sebuah proses menganalisa data untuk mengetahui suatu pola dari kelompok data yang tersembunyi (Mardi, 2019). Yang bertujuan untuk menemukan, menggali pengetahuan dari data atau informasi yang kita miliki. *Data mining* melibatkan pencarian pola yang diperlukan dalam database untuk membantu pembuat keputusan di masa yang akan datang.

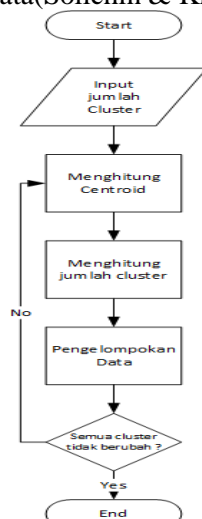


Gambar 1 Teknik Data mining

### 2.3 Metode K Means Clustering

Metode *K-Means* pertama kali di perkenalkan oleh MacQueen JB pada tahun 1976. Metode ini adalah salah satu metode *non hierarchi* yang umum digunakan. Metode termasuk dalam teknik penyekatan (*partition*) yang membagi atau memisahkan objek *k* ke daerah bagian yang terpisah (Sibuea & Sapta, 2017). *K-Means* merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi database kedalam beberapa clasteran *k*. Algoritma cukup mudah untuk diimplementasikan dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan. Prinsip utama dari teknik ini adalah menyusun *K* buah partisi/pusat massa (centroid)/rata-rata (mean) dari sekumpulan data. Algoritma *K-Means* dimulai dengan pembentukan partisi kluster diawal kemudian secara iteraktif partisi claster ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi claster (Sibuea et al., 2017). Algoritma klasterisasi *K-Means* merupakan salah satu dari algoritma data mining yang populer di kalangan peneliti. Jika dibandingkan dengan algoritma klasterisasi lainnya, *k-means* juga

masih cukup tangguh di berbagai jenis data (Solichin & Khairunnisa, 2020).



Gambar 2 Flowchart Metode K-Means

## 2.4 Analisis

Analisis adalah langkah-langkah dalam penyelesaian suatu masalah, dalam metode *clustering* algoritma *K-Means*, algoritma *K-Means* merupakan model *centroid* atau titik tengah yang menggunakan *centroid* untuk membuat *cluster*. Maka dilakukan tahapan-tahapan sebagai berikut :

1. Penetapan jumlah *cluster* (K).
2. Hitung jarak setiap data ke pusat *cluster* antara objek ke *centroid*.
3. Kelompokkan data ke dalam *cluster* dengan jarak minimal.
4. Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian di hitung kembali pusat *cluster* yang baru berdasarkan data penggabungan covid provinsi dengan menggunakan *K-Means*.

**Tabel 1 Normalisasi Data Yang Digunakan Pada Periode 2021**

No	Nama Provinsi	Kasus Meninggal	Kasus Positif
1	Aceh	38.416	2.066
2	Bali	114.233	4.046
3	Banten	132.693	2.688
4	Bengkulu	23.104	473
5	Daerah Istimewa Yogyakarta	156.769	5.263
6	DKI Jakarta	864.045	13.596
7	Gorontalo	11.834	460
8	Jambi	29.768	789
9	Jawa Barat	707.934	14.737
10	Jawa Tengah	486.435	30.225
11	Jawa Timur	399.478	29.697
12	Kalimantan Barat	41.536	1.061
13	Kalimantan Selatan	69.927	2.390
14	Kalimantan Tengah	45.622	1.430
15	Kalimantan Timur	158.245	5.449
16	Kalimantan Utara	35.885	811
17	Kepulauan Bangka Belitung	52.290	1.455
18	Kepulauan Riau	53.884	1.748
19	Lampung	49.691	3.866
20	Maluku	14.587	261
21	Maluku Utara	12.101	303
22	Nusa Tenggara Barat	27.777	815

23	Nusa Tenggara Timur	64.035	1.340
24	Papua	34.341	559
25	Papua Barat	23.378	357
26	Riau	128.825	4.109
27	Sulawesi Barat	12.353	346
28	Sulawesi Selatan	109.919	2.238
29	Sulawesi Tengah	47.153	1.602
30	Sulawesi Tenggara	20.160	527
31	Sulawesi Utara	34.715	1.056
32	Sumatera Barat	89.849	2.152
33	Sumatera Selatan	59.937	3.071
34	Sumatera Utara	106.045	2.889

Berikut ini langkah-langkah pada algoritma *K-Means* sampai diketahui pembagian nilai *centroid* sebelumnya tidak berubah.

1. Menentukan jumlah *cluster* misalkan sebanyak  $k = 3$
2. Menentukan *centroid*  $c$  setiap *cluster* yang diambil dari data sumber

**Table 2 Tabel Data Centroid Awal**

Centroid	Data	Kasus Positif	Kasus Meninggal
Centroid 1	DKI Jakarta	864.045	13.596
Centroid 2	Kepulauan Bangka Belitung	52.290	1.455
Centroid 3	Gorontalo	11.834	460

Hitung jarak data ke *Centroid* menggunakan rumus *Euclidean*, data tersebut dari *Cluster* terdekatnya.

- a. Jarak antara Provinsi nomor pertama dengan titik  $m_1$
- b. Jarak antara Provinsi nomor pertama dengan titik  $m_2$

Untuk lebih lengkapnya jarak pada setiap baris data, hasilnya seperti pada tabel 3 berikut:

**Table 3 Tabel Hasil Perhitungan Iterasi Ke 1**

No	Nama Provinsi	C1	C2	C3	Jarak Terdekat	Nilai
1	Aceh	825709,505	13887,447	26630,471	C2	13887,447
2	Bali	749872,814	61997,165	102461,771	C2	61997,165
3	Banten	731433,341	80412,454	120879,535	C2	80412,454
4	Bengkulu	841043,387	29202,516	11270,007	C3	11270,007
5	Daerah Istimewa Yogyakarta	707325,087	104548,373	145014,561	C2	104548,373
33	Sumatera Selatan	804176,878	7815,885	48173,810	C2	7815,885
34	Sumatera Utara	758075,616	53774,124	94242,308	C2	53774,124
Total WCV						1660408,420

Menghitung nilai BCV dari 3 *centroid* yang digunakan dapat dilihat pada tabel 4 berikut.

**Table 4 Data Nilai Cluster Rata - Rata**

Centroid	Nilai 1	Nilai 2
M1	864.045	13.596
M2	417	5685
M3	108	142

**Table 5 Tabel Hasil Perhitungan Iterasi Ke 2**

No	Nama Provinsi	C1	C2	C3	Jarak	Nilai
----	---------------	----	----	----	-------	-------

						Terdekat
1	Aceh	647957,10	78020,86	10198,35	C3	10.198,4
2	Bali	572114,28	2292,07	85998,38	C2	2.292,1
3	Banten	553700,88	16444,67	104414,31	C2	16.444,7
4	Bengkulu	663307,51	93385,54	5202,37	C3	5.202,4
5	Daerah Istimewa Yogyakarta	529560,93	40383,36	128550,62	C2	40.383,4
33	Sumatera Selatan	626416,99	56478,30	31722,25	C3	31.722,3
34	Sumatera Utara	580331,33	10522,95	77777,24	C2	10.522,9
Total WCV						1655096873 67,21

**Tabel 6 Hasil Modal Error**

No	Nama Provinsi	Hasil Sum Square Error	Selisih
1	C1	877641	171983,7
2	C2	53745	67469,7
3	C3	12294	6808,9

Dari Hasil Modal Error, dilakukan sebanyak 2 iterasi maka nilai *centroid*nya tidak ada perubahan lagi maka hasil adalah sebagai berikut:

$$BCV/WCV = 1316231,89 / 165509687367,21 \\ = 7.9525$$

Hasil pengelompokan *cluster* dari Provinsi yang berzona covid adalah sebagai berikut:

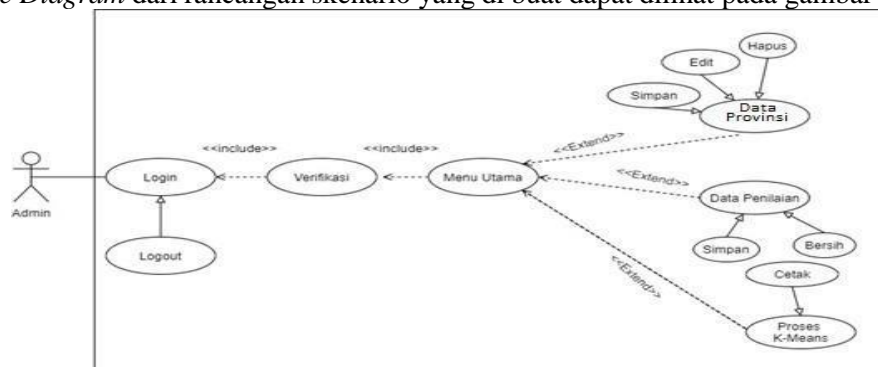
**Table 7 Hasil Pengelompokan**

No	Pengelompokan	Kode Wilayah
1	Zona Merah	6, 9, 10, 11
2	Zona Kuning	2, 3, 5, 13, 15, 23, 26, 28, 32, 33, 34
3	Zona Hijau	1, 4, 7, 8, 12, 14, 16, 17, 18, 19, 20, 21, 22, 24, 25, 27, 29, 30, 31

## 2.5 Perancangan Sistem

### a. Use Case Diagram

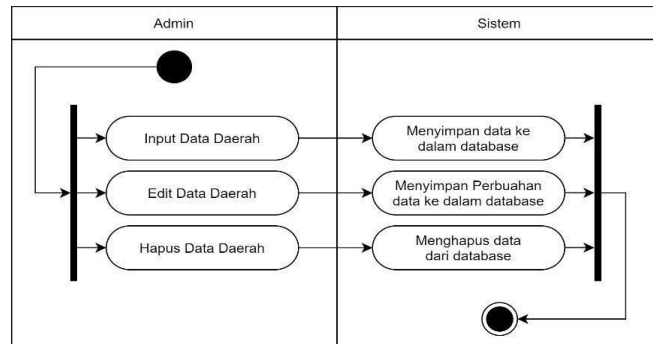
Use Case Diagram dari rancangan skenario yang di buat dapat dilihat pada gambar 3.



Gambar 3 Use Case Diagram Form Proses Metode K-Means

### b. Activity Diagram Data Daerah

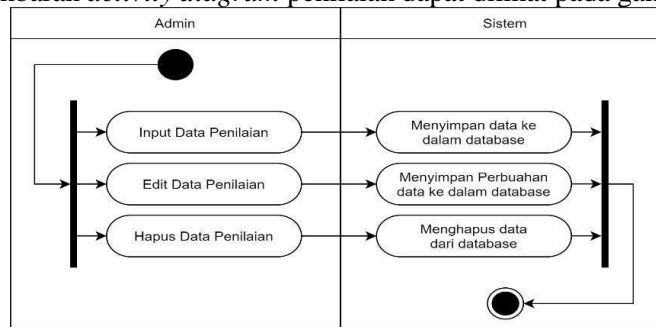
Adapun gambaran *activity diagram* data daerah dapat dilihat pada gambar 4.



Gambar 5 Activity Diagram Data Daerah

c. Activity Diagram Penilaian

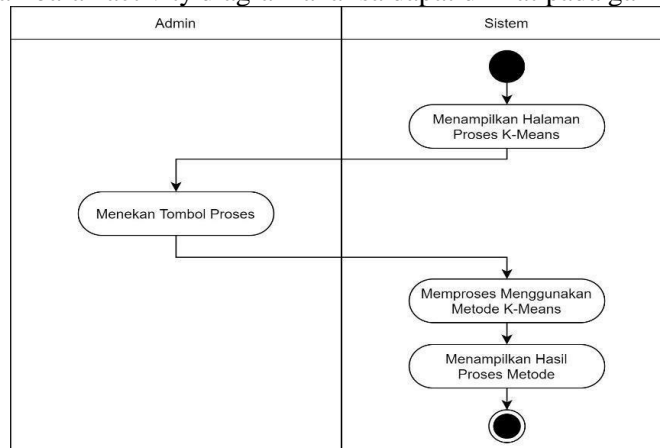
Adapun gambaran activity diagram penilaian dapat dilihat pada gambar 5.



Gambar 5 Activity Diagram Penilaian

d. Activity Diagram Analisa

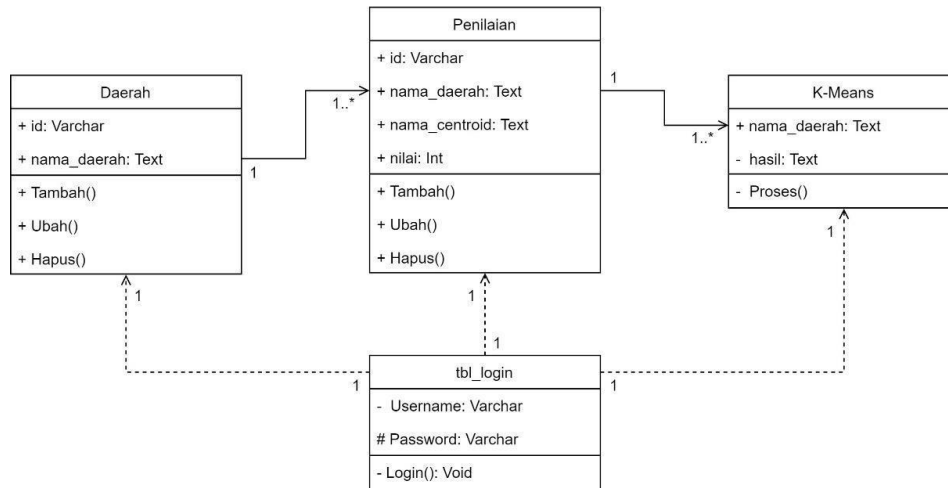
Adapun gambaran activity diagram analisa dapat dilihat pada gambar 6.



Gambar 6 Activity Diagram Analisa

e. Pemodelan Class Diagram

Berdasarkan deskripsi dari use case diagram diatas berikut ini adalah gambar class diagram yaitu sebagai berikut:

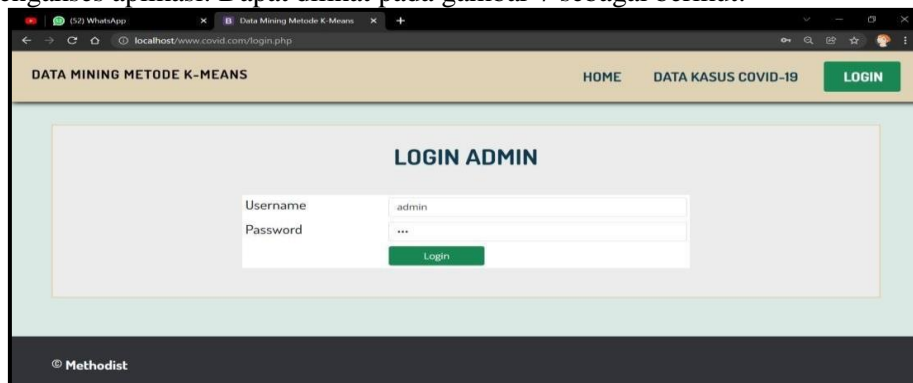


Gambar 6 Class Diagram

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Tampilan Halaman Login

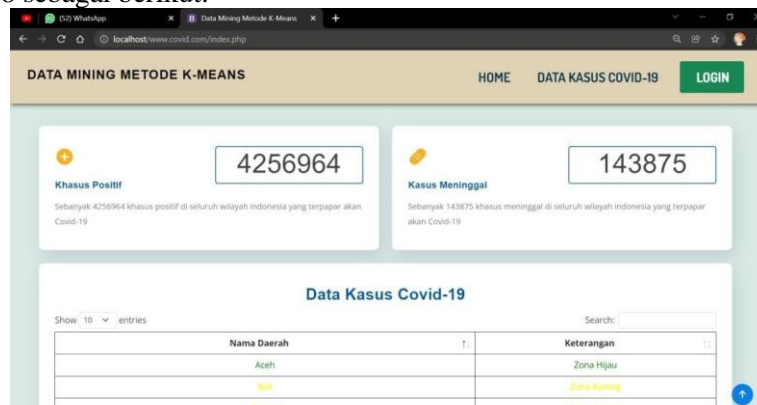
Tampilan halaman login merupakan tampilan yang pertama kali ditemukan saat membuka *website*. Halaman ini ditujukan untuk admin memasukkan username dan kata sandi yang benar agar bisa mengakses aplikasi. Dapat dilihat pada gambar 7 sebagai berikut:



Gambar 7 Tampilan Halaman Login

#### 3.2 Tampilan Halaman utama

Tampilan halaman utama merupakan tampilan awal aplikasi. Halaman ini berisikan artikel penjelasan tentang covid-19 dan data kasus covid-19 pada tiap daerah. Dapat dilihat pada gambar 8 sebagai berikut.



Gambar 8 Tampilan Halaman Utama

#### 3.3 Tampilan Data Daerah

Tampilan Data Daerah merupakan tampilan yang menampilkan data daerah yang diinput. Dapat dilihat pada gambar 9 sebagai berikut.

No	Iso Code	Nama Daerah	
1	ID-BA	Aceh	[Add] [Edit] [Del]
2	ID-BL	Bali	[Add] [Edit] [Del]
3	ID-BT	Banten	[Add] [Edit] [Del]
4	ID-BE	Bengkulu	[Add] [Edit] [Del]
5	ID-YO	Daerah Istimewa Yogyakarta	[Add] [Edit] [Del]
6	ID-JK	DKI Jakarta	[Add] [Edit] [Del]
7	ID-GO	Gorontalo	[Add] [Edit] [Del]

Gambar 9 Tampilan Data Daerah

### 3.4 Tampilan Analisa K Means

Tampilan analisa K Means merupakan tampilan yang menampilkan data proses K means. Dapat dilihat pada gambar 10 berikut.

nama	centroid	
	Kasus Meninggal	Kasus Positif
Aceh	38416	2066
Bali	114233	4046
Banten	132693	2688
Bengkulu	23104	473
Daerah Istimewa Yogyakarta	156769	5263
DKI Jakarta	864045	13596

Gambar 10 Tampilan Proses K-Means

## 4. KESIMPULAN DAN SARAN

### 4.1 Kesimpulan

Berdasarkan hasil analisa dari permasalahan yang terjadi dengan kasus yang dibahas pengelompokan daerah zona covid-19 dengan menerapkan sistem pengelompokan covid terhadap sistem yang di rancang dan di bangun maka dapat di tarik kesimpulan sebagai berikut:

1. Dengan mengimplementasikan aplikasi dengan sistem pengelompokan covid-19 dapat mengetahui zona di setiap provinsi di indonesia apakah termasuk zona hijau, zona kuning, zona merah dan akan dapat diakses oleh pengguna maupun admin.
2. Penerapan metode K-means clustering dilakukan dengan mengolah data dari setiap provinsi di Indonesia dimana zona merah terdiri dari 4 provinsi, zona kuning terdiri dari 11 provinsi, dan zona hijau terdiri dari 19 provinsi.

### 4.2 Saran

Untuk meningkatkan kemampuan dan fungsi dari sistem ada beberapa saran yang dapat diberikan untuk pengembangan yang bisa dilakukan yaitu:

1. Sistem dapat dikembangkan dalam menggunakan bahasa pemrograman yang lebih canggih seperti bahasa pemrograman android ataupun dengan penggunaan aplikasi dalam bentuk *mobile*.
2. Disarankan dengan menggunakan sistem yang lain dari sistem pengelompokan covid sebagai perbandingan hasil yang didapat untuk meningkatkan penentuan zona lokasi dengan cepat dan tepat.
3. Diharapkan dapat melakukan pengembangan dan penyempurnaan terhadap sistem dan dibuat sistem *user interface* yang lebih menarik.



## REFERENSI

- [1] Nur, A., Thohari, A., & Vernandez, A. B. (2020). Aplikasi Monitoring Kasus Coronavirus Berbasis Android. *JTET (Jurnal Teknik Elektro Terapan) Polines*, 9(1), 12–17.
- [2] Sibuea, F. L., & Sapta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 85 - 92.
- [3] Sindi, S., Ningse, W. R. O., Sihombing, I. A., Ilmi R.H.Zer, F., & Hartama, D. (2020). Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia. *Jti (Jurnal Teknologi Informasi)*, 4(1), 166–173.
- [4] Solichin, A., & Khairunnisa, K. (2020). Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means. *Fountain of Informatics Journal*, 5(2), 52. <https://doi.org/10.21111/fij.v5i2.4905>
- [5] Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine and International Health*, 25(3), 278–280. <https://doi.org/10.1111/tmi.13383>
- [6] Dan, T. T., Sihwi, S. W., & Anggrainingsih, R. (2015). Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret. *JURNAL ITSMART*, 84-91.
- [7] Hendini, A. (2016). Pemodelan UML Sistem Informasi Monitoring Penjualan Dan Stok Barang (Studi Kasus: Distro Zhezha Pontianak). *JURNAL KHATULISTIWA INFORMATIKA*, 107- 116.
- [8] Indraputra, R. A., & Fitriana, R. (2017). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 275-282.
- [9] Mardi, Y. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 213-219.
- [10] Merliana, N. P., Ernawati, & Santoso, A. J. (2018). Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering. *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI\_U)*, 1-7.
- [11] Novianto, D. (2015). Pemfilteran Hypertext Transfer Protocol Secure Untuk Penggunaan Internet Yang Aman. *Jurnal Informatika*, 24-32.
- [12] Riyanto, B. (2019). Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan). *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 562-568.
- [13] Sujarwo, A., Iswanto, A. C., & Nugroho, K. (2019). Sistem Penilaian Kinerja Pegawai Dalam Pelayanan Nasabah Pada BPR Agung Sejahtera Menggunakan Unified Modeling Language. *Fakultas Teknik Universitas Wahid Hasyim*, 270-275.
- [14] Susilo, A., Sinto, R., Santoso, W. D., & Mira, Y. (2020). Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesi*, 45-67.
- [15] Syafitri, Y. (2019). Analisa Dan Perancangan Berbasis UML Pada Sistem Informasi Simpan Pinjam Koperasi Swamitra Bandar Lampung. *Dian Cipta Cendikia*, 22-31.
- [16] Wakhidah, N. (2018). Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering). *Jurnal Informatika*, 1-8.