



### Computer Based Information System Journal

ISSN (Print): 2337-8794 | E- ISSN : 2621-5292  
 web jurnal : <http://ejournal.upbatam.ac.id/index.php/cbis>



## Implementasi Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Sumatera Barat

**Dian Permata Sari**

STMIK Jayanusa, Indonesia.

#### INFORMASI ARTIKEL

Diterima Redaksi: 6 Februari 2021  
 Diterbitkan Online: 24 Maret 2021

#### KATA KUNCI

Data Mining, covid-19, K-Means

#### KORESPONDENSI

E-mail:  
 ummuasheeqa07@gmail.com

#### A B S T R A C T

*Corona viruses (CoV) are part of a family of viruses that cause illnesses ranging from the flu to more severe illnesses such as Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). This virus can be transmitted from person to person through small droplets from the nose or mouth when coughing, sneezing or speaking. Because the spread of this virus is very fast, it requires fast handling so that this virus does not spread, one of which is by implementing health protocols, namely maintaining distance, washing hands and using masks. All provinces in Indonesia have not been spared from this virus, including the province of West Sumatra. Classification of the spread of this virus is necessary in order to break the chain of its spread. One of the techniques used in this grouping is k-means, which uses several groups to assign multiple data to a partition system. The results of this study indicate that the regions in the first cluster have the highest rates of positive cases and patients who die, while the areas in the second and third clusters have the potential for the spread of Covid-19 which is also a concern of the government.*

### I. Latar Belakang

Dipenghujung tahun 2019 dunia dihebohkan dengan penemuan virus jenis baru yang sangat mematikan. Penyakit ini teridentifikasi pertama kali di daerah wuhan provinsi hubei, Tiongkok.

Pada awalnya virus ini diduga akibat paparan pasar grosir makanan laut wuhan yang banyak menjual spesies hewan hidup. Penyakit ini dengan cepat menyebar di dalam negeri ke bagian lain China . [1]. Indoensia pada awalnya masuk dalam salah satu Negara yang belum terinfeksi virus corona akhirnya pada awal maret

2020 mengumumkan kasus pertama pasien terinfeksi virus corona sebanyak dua orang. Dan kemudian kasus tersebut bertambah setiap harinya hingga pada bulan januari 2021 sudah tembus lebih dari satu juta kasus terkonfirmasi positif, dan Indonesia termasuk Negara yang tingkat kematiannya tinggi akibat virus corona.

Penyakit yang disebabkan virus corona, atau dikenal dengan COVID19, adalah jenis baru yang ditemukan pada tahun 2019 dan belum pernah diidentifikasi menyerang manusia sebelumnya.[2] Virus ini menyebabkan infeksi

saluran pernapasan atas ringan hingga sedang, seperti penyakit flu.

Munculnya 2019-nCoV telah menarik perhatian global, dan Pada 30 Januari WHO telah menyatakan COVID-19 sebagai darurat kesehatan masyarakat yang menjadi perhatian internasional[1].

Penyebaran kasus COVID-19 yang merata di seluruh provinsi di Indonesia, merupakan penyebaran yang cukup cepat dan berdampak negatif pada seluruh bidang.[3] Salah satu provinsi yang tak luput dari serangan covid-19 adalah Sumatera Barat. Seluruh kota dan kabupaten yang ada terpapar covid-19 dengan jumlah kasus berbeda-beda tiap daerahnya. Tujuan dilakukannya penelitian ini untuk mengelompokkan persebaran Virus Corona di provinsi Sumatera Barat berdasarkan parameter jumlah kasus positif dan pasien yang meninggal.

Penelitian ini menggunakan metode k-means dalam mengelompokkan sebaran kasus covid-19 di daerah Sumatera Barat. K-Means merupakan salah satu algoritma Clustering yang masuk dalam kelompok *Unsupervised learning* yang digunakan untuk mengelompokkan data kedalam beberapa kelompok dengan sistem partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma K-Means, komputer menerima data-data yang tidak diketahui kelasnya terlebih dahulu lalu mengelompokkannya. Input yang diterima ialah data dan jumlah kelompok (cluster) yang diinginkan.[3]

Beberapa penelitian memanfaatkan algoritma klusterisasi k-means untuk berbagai keperluan, antara lain pengelompokkan jumlah kematian penderita covid-19 berdasarkan negara di benua Asia.[4], menentukan tingkat penyebaran pandemi covid-19 di Indonesia. [3] penyebaran penyakit diare di DKI Jakarta[5], Klusterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta [6]

## II. Kajian Literatur

<http://ejournal.upbatam.ac.id/index.php/cbis>

### A. Data mining

Data mining merupakan salah satu dari rangkaian adalah serangkaian *Knowledge Discovery Database (KDD)*. [7] (KDD) berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahap yaitu (Tan, 2004) : pembersihan data, integrasi data, transformasi data, aplikasi teknik data mining, evaluasi pola yang ditemukan, presentasi pengetahuan.

### B. Covid-19

*Coronavirus* adalah virus RNA dengan ukuran partikel 120-160 nm. Virus ini utamanya menginfeksi hewan, termasuk di antaranya adalah kelelawar dan unta.[8]

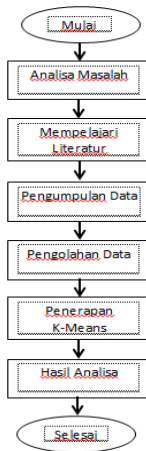
Virus ini memiliki sifat dapat menularkan antara hewan dan manusia seperti penyakit rabies dan malaria. Nama virus ini diambil dari bahasa latin yang berarti mahkota karena bentuk dari virus itu sendiri memiliki paku yang menonjol menyerupai mahkota dan korona matahari.

Penyebaran virus ini sangat cepat, adapun cara penyebarannya yaitu melalui percikan air liur (droplet), melalui udara, permukaan yang terkontaminasi, melalui limbah manusia. Selain itu ada tempat-tempat yang menjadi rawan penyebaran virus ini yaitu tempat yang ramai, tempay yang sempit dan ruangan yang terbatas dan tertutup.

## III. Metodologi

### 3.1 Tahap Pengumpulan Data

Dalam metode ini cara untuk memperoleh dan mengumpulkan data-data ilmiah yang dibutuhkan dengan fungsi dan tujuan tertentu. Metode penelitian dapat dilihat pada gambar di bawah ini:



Gambar 1. Metode Penelitian

### 3.2 Algoritma K-Means

Clustering juga disebut sebagai *segmentation*. Metode ini digunakan untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada sebuah kelompok atribut, mengelompokkan data yang memiliki kemiripan atribut.

*K-Means* merupakan *algoritma clustering* yang berulang-ulang. *Algoritma K-Means* menetapkan nilai-nilai *cluster* (k) secara *random*, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, *mean* atau “*means*”. Kemudian menghitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus *Euclidian* hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Klasifikasikan setiap data berdasarkan kedekatannya dengan *centroid*. Lakukan langkah tersebut hingga nilai *centroid* tidak berubah (stabil). [7]

Berikut adalah tahapan-tahapan untuk melakukan optimasi menggunakan *algoritma K-Means* :

- a. Tentukan pusat cluster secara acak
- b. Hitung jarak setiap data yang ada terhadap setiap cluster.

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2} \tag{1}$$

Keterangan:

(x ,y) = Koordinat objek

(s ,t) = Koordinat *Centroid*

i = Banyaknya objek

- c. Suatu data akan menjadi anggota baru dari suatu cluster yang memiliki jarak terkecil dari pusat cluster-nya.

- d. Hitung pusat cluster baru

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \tag{2}$$

Keterangan:

$V_{ij}$  = *Centroid* rata – rata pada *Cluster* ke – i untuk variabel ke –j

$N_i$  = Jumlah anggota *Cluster* ke –i

i,k = Indeks dari *Cluster*

j = Indeks variabel

$X_{kj}$  = Nilai data ke –k variabel ke – j untuk *Cluster* tersebut

- e. Ulangi langkah 2 hingga posisi data sudah tidak mengalami perubahan.

## IV. Pembahasan

### A. Perhitungan manual dengan Algoritma *K-Means*

Sampel yang digunakan dalam penelitian ini adalah 19 kota dan kabupaten yang ada di provinsi sumatera barat yang terkena dampak penyebaran COVID-19. Berikut adalah sampel data penyebaran COVID-19 di sumatera barat per tanggal 21 Januari 2021:

Tabel 1. Sampel Data Sebaran Covid-19

No	Kabupaten/Kota	Kasus Positif	Kasus Meninggal
1	Padang	13.233	269
2	pariaman	537	16
3	Kab padang pariaman	1045	39
4	Padang panjang	789	16
5	Kab. Sijunjung	751	19
6	Bukittinggi	1009	18
7	Kab. Agam	1760	31
8	Kab. Dharmasraya	511	7
9	Kab. Solok	736	18
10	Kab. Pasaman barat	482	31
11	Kab. Solol selatan	332	4
12	Kab. Pesisir selatan	1061	25
13	Payakumbuh	701	9
14	Kota Solok	573	14
15	Kab. Lima puluh kota	459	15
16	Sawahlunto	380	6

17	Kab. Kep. Mentawai	349	0
18	Kab. Tanah datar	1051	34
19	Kab. Pasaman	309	10

Tabel diatas merupakan data sebaran covid-19 di seluruh kota dan kabupaten yang ada di Sumatera barat per tanggal 21 Januari 2021. Terdapat 2 atribut dari data di atas yaitu kasus positif dan kasus meninggal.

Iterasi ke-1

1. Menentukan pusat *cluster* secara acak.

C1 (Padang) = (13233,269), C2 (Kabupaten Agam) = (1760,31), C3 (Kabupaten Pasaman) = (309,10).

2. Menghitung jarak setiap data yang ada terhadap setiap pusat *cluster*.

Pada tahap ini jarak terdekat antara data dengan *Cluster* akan menentukan suatu data masuk ke dalam *Cluster* mana. Di bawah ini adalah contoh perhitungan iterasi 1

$$\text{Padang (C1)} = \sqrt{(13233 - 13233)^2 + (269 - 269)^2} = 0$$

$$\text{Padang (C2)} = \sqrt{(13233 - 1760)^2 + (269 - 31)^2} = 11475,47$$

$$\text{Padang (C3)} = \sqrt{(13233 - 309)^2 + (269 - 10)^2} = 12926,59$$

Berikut adalah hasil seluruh perhitungan jarak pada Iterasi ke 1.

Tabel 2. Hasil Perhitungan Iterasi 1

Kota/Kabupaten	Jarak Ke Centroid		
	C1	C2	C3
Padang	0	11475,47	12926,59
pariaman	12689,52	1223,092	228,0789
Kab padang pariaman	12190,17	715,0448	736,5711
Padang panjang	12446,57	971,1159	480,0375
Kab. Sijunjung	12484,5	1009,071	442,0916
Bukittinggi	12226,58	751,1125	700,0457
Kab. Agam	11475,47	0	1451,152
Kab. Dharmasraya	12724,7	1249,231	202,0223
Kab. Solok	12499,52	1024,083	427,0749
Kab. Pasaman barat	12753,22	1278	174,2699
Kab. Solol selatan	12903,72	1428,255	23,76973
Kab. Pesisir selatan	12174,45	699,0258	752,1496
Payakumbuh	12534,7	1059,228	392,0013
Kota Solok	12662,57	1187,122	264,0303

Kab. Lima puluh kota	12776,53	1301,098	150,0833
Sawahlunto	12855,69	1380,226	71,11259
Kab. Kep. Mentawai	12886,81	1411,34	41,23106
Kab. Tanah datar	12184,27	709,0063	742,388
Kab. Pasaman	12926,59	1451,152	0

Pada tabel 2 di atas merupakan hasil semua perhitungan jarak pada iterasi 1 setelah itu akan dilakukan pengelompokan berdasarkan *cluster*.

3. Melakukan pengelompokan berdasarkan *cluster*.

Suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terkecil dari pusat *cluster* nya. Misalkan untuk data pertama, jarak terkecil diperoleh pada *cluster* pertama, sehingga data pertama akan menjadi anggota dari *cluster* pertama. Posisi *cluster* selengkapnya dapat dilihat pada tabel 3 di bawah ini:

Tabel 3. Pengelompokan *Cluster* Iterasi 1

Kota/Kabupaten	Jarak Terdekat
Padang	C1
pariaman	C3
Kab padang pariaman	C2
Padang panjang	C3
Kab. Sijunjung	C3
Bukittinggi	C3
Kab. Agam	C2
Kab. Dharmasraya	C3
Kab. Solok	C3
Kab. Pasaman barat	C3
Kab. Solol selatan	C3
Kab. Pesisir selatan	C2
Payakumbuh	C3
Kota Solok	C3
Kab. Lima puluh kota	C3
Sawahlunto	C3
Kab. Kep. Mentawai	C3
Kab. Tanah datar	C2
Kab. Pasaman	C3

Dari hasil pengelompokan iterasi 1 dapat disimpulkan bahwa Padang merupakan satu-

satunya daerah yang berada pada cluster 1 (C1), untuk C2 ada empat daerah yaitu kab. Padang pariaman, kab. Agam, kab. Pesisir selatan dan kab. Tanah datar da nada 14 daerah masuk dalam cluster 3 (C3).

4. Menentukan *Centroid* baru untuk iterasi berikutnya dengan mencari rata-rata nilai dari setiap *cluster*.

Berikut adalah nilai *centroid* baru untuk iterasi 2:

centroid	c1	13233	269
	c2	1229,25	32,25
	c3	565,5714	13,07143

Perhitungan akan dihentikan jika tahapan iterasi telah mencapai hasil yang sama tanpa ada perpindahan lagi.

Pada penelitian ini iterasi dilakukan sampai iterasi ketiga. Hasil akhir posisi *Cluster* dari *Centroid* iterasi ketiga dapat dilihat pada tabel 4 dibawah ini:

Tabel 4. Pengelompokkan *Cluster* Iterasi ke-3

Kota/Kabupaten	Jarak Terdekat
Padang	C1
pariaman	C3
Kab padang pariaman	C2
Padang panjang	C3
Kab. Sijunjung	C3
Bukittinggi	C2
Kab. Agam	C2
Kab. Dharnasraya	C3
Kab. Solok	C3
Kab. Pasaman barat	C3
Kab. Solol selatan	C3
Kab. Pesisir selatan	C2
Payakumbuh	C3
Kota Solok	C3
Kab. Lima puluh kota	C3
Sawahlunto	C3
Kab. Kep. Mentawai	C3
Kab. Tanah datar	C2
Kab. Pasaman	C3

a. Implementasi pada *RapidMiner*

<http://ejournal.upbatam.ac.id/index.php/cbis>

Pada Implementasi dan Pengujian penulis menggunakan aplikasi *RapidMiner Studio* Versi 5.3. Tujuannya untuk membandingkan hasil perhitungan secara manual metode *K-Means* dengan pengujian data menggunakan aplikasi *RapidMiner*.

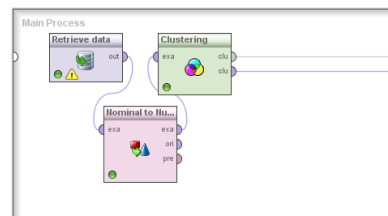
Berikut adalah langkah-langkah mengaplikasikan *rapidminer* untuk metode *K-Means*:

1. *Import* data kedalam *RapidMiner* dalam bentuk sheet Excel dan tentukan tipe datanya

no	kab/kota	kasus positif	kasus menul
1	Padang	13233	269
2	pariaman	537	16
3	Kab padang	1045	39
4	Padang pan,	789	16
5	Kab. Sijunju	751	19
6	Bukittinggi	1009	18
7	Kab. Agam	1760	31
8	Kab. Dharna	511	7
9	Kab. Solok	736	18
10	Kab. Pasam	482	31

Gambar 2. Penyesuaian Atribut dan Tipe data Sebelum data diimport maka data sudah diketik terlebih dahulu di Microsoft excel. Kemudian buka aplikasi *RapidMiner*. Setelah data diimport kemudian tentukan tipe data yang digunakan misal jika data berupa nama maka pilih tipe data text, untuk angka gunakan tipe data integer.

2. Menghubungkan data dengan model *K-Means*

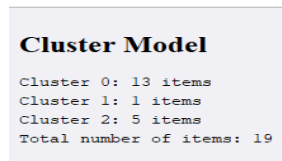


Gambar 3. Menghubungkan data dengan model *K-Means*

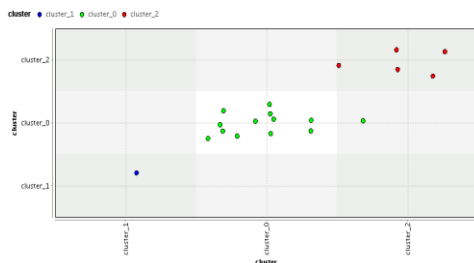
Setelah data berhasil di import maka data di drag ke panel proses kemudian data akan dihubungkan dengan k-means yang sebelumnya telah dicari pada panel operator dan setelah itu project dapat dijalankan.

3. Jalankan *project*

Tampilan *text view*



Gambar 4. Hasil Tampilan *Text View*  
Tampilan *plot view*



Gambar 5. Hasil Tampilan *Plot View*

Warna biru pada gambar di atas menunjukkan cluster 1 warna hijau *cluster* 0 sedangkan warna merah cluster 2.

Setelah melakukan implementasi menggunakan Aplikasi *RapidMiner*, hasil yang didapat setelah di implementasikan pada *RapidMiner* sama dengan hasil pada perhitungan manual.

## V. KESIMPULAN

Berdasarkan analisa dan hasil perhitungan menggunakan Algoritma K-Means dalam penyebaran COVID-19 di provinsi Sumatera Barat, dapat ditarik kesimpulan sebagai berikut :

1. Daerah yang menjadi pusat *cluster* pertama adalah kota Padang dan ini merupakan daerah penyebaran covid-19 terbesar di provinsi Sumatera Barat dengan tingkat kasus positif dan meninggal paling besar.
2. Daerah-daerah yang berada pada *cluster* kedua dan ketiga cukup berpotensi untuk penyebaran covid-19 di Sumatera Barat
3. Pada penelitian selanjutnya, *clusterisasi* dapat dikembangkan dengan menerapkan berbagai metode *clusterisasi* lainnya serta dengan jumlah variabel yang lebih banyak.

## DAFTAR PUSTAKA

- [1] R. N. Putri, "Indonesia dalam Menghadapi Pandemi Covid-19," *J. Ilm. Univ. Batanghari Jambi*, vol. 20, no. 2, p. 705, 2020, doi: 10.33087/jiubj.v20i2.1010.
- [2] A. F. B. Watratan, A. Puspita, and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol. ( Jacost )*, vol. 1, no. 1, pp. 7–14, 2020.
- [3] P. P. P. A. N. . F. I. R. . Z. Nayuni Dwitri , Jose A Tampubolon , Sandi Prayoga and D. Hartama, "PENERAPAN ALGORITMA K-MEANS DALAM MENENTUKAN TINGKAT PENYEBARAN PANDEMI COVID-19 DI INDONESIA," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 128–132, 2020.
- [4] Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradig. Inform. dan Komput.*, vol. 22, no. 2, pp. 183–188, 2020.
- [5] T. B. Santos, "Aplikasi Data Mining untuk Clustering Daerah Penyebaran Penyakit Diare di DKI Jakarta Menggunakan Algoritma K-MEANS," *J. Ilm. FIFO*, vol. 11, no. 2, p. 131, 2019, doi: 10.22441/fifo.2019.v11i2.003.
- [6] A. Solichin and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means," *Fountain Informatics J.*, vol. 5, no. 2, p. 52, 2020, doi: 10.21111/fij.v5i2.4905.
- [7] R. T. Vulandari, *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media, 2017.
- [8] A. Susilo *et al.*, "Coronavirus Disease 2019: Tinjauan Literatur Terkini," *J. Penyakit Dalam Indones.*, vol. 7, no. 1, p. 45, 2020, doi: 10.7454/jpdi.v7i1.415.
- [9] D. N. Anjar Wanto, Muhammad Noor Hasan Siregar, Agus Perdana Windarto, Dedy Hartama, Ni Luh Wiwik Sri Rahayu Ginantra, M. R. L. S. V. D. Edi

Surya Negara, and C. Prianto, *Data Mining: Algoritma dan Implementasi*. Medan: Yayasan Kita Menulis, 2020.