

PENGELOMPOKAN MASYARAKAT DI WILAYAH KABUPATEN SIKKA YANG TERDAMPAK COVID MENGGUNAKAN ALGORITMA K-MEANS

Febriyanti Alwisye Wara¹, Cornelia Hildegardis², Imelda Dua Reja³

^{1,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Nusa Nipa

² Program Studi Teknik Arsitektur, Fakultas Teknik, Universitas Nusa

Nipa e-mail: Fwara840@gmail.com¹, childegardis4@gmail.com²

ABSTRAK

Pandemic yang disebabkan oleh covid-19 telah tersebar di seluruh wilayah Indonesia, salah satunya adalah Kabupaten Sikka yang berada di Provinsi NTT. Ada 21 kecamatan yang berada di kabupaten Sikka dan tiap harinya angka kejadian kasus covid-19 terus bertambah. Tujuan penelitian ini adalah mengetahui bagaimana mengelompokkan masyarakat yang terdampak covid menggunakan Algoritma K-Means, dimana dari data yang didapat akan ada 3 centroid pusat kemudian dilakukan iterasi pada data tersebut, sehingga nantinya dapat diketahui kecamatan mana saja yang memiliki kasus terbanyak, sedang dan rendah. Dalam penelitian ini Jenis pengelompokan yang diambil yaitu berdasarkan status yang terdiri dari kasus sembuh dan meninggal perkecamatan, dari data kasus yang didapat selanjutnya akan di buat pengelompokan dengan Algoritma K-Means. Setelah dilakukan pengujian terhadap data kasus covid-19 di wilayah kabupaten Sikka maka didapat kelompok kecataman terbanyak yang terdampak ada 2 kecamatan, cluster sedang terdapat 6 Kecamatan dan cluster rendah terdapat 13 kecamatan.

Kata Kunci: pengelompokan masyarakat, algoritma K-Means, dan Pandemi Covid 19

1. PENDAHULUAN

Dewasa ini, dunia sedang diguncang oleh pandemik hebat bernama Covid-19 (Corona Virus Disease). Peningkatan dari hari kehari jumlah pasien terinfeksi virus Covid-19 sudah sulit dikendalikan diperlukannya suatu perencanaan yang jelas dan lugas dari pemerintah untuk menanggulangi permasalahan ini [1]. Virus Covid-19 merupakan penyakit menular yang ditandai oleh gejala pada bagian pernapasan akut (Coronavirus 2 (Severe Acute Respiratory Syndrome Coronavirus 2) atau SARS-CoV-2) [2]. Infeksi Virus Corona atau Covid-19 disebabkan oleh Corona virus, yaitu kelompok virus yang menginfeksi sistem pernapasan, pada sebagian besar kasus corona virus hanya menyebabkan infeksi pernapasan ringan sampai sedang, seperti flu, akan tetapi, virus ini juga bisa menyebabkan infeksi pernapasan berat, seperti Pneumonia, MiddleEast Respiratory Syndrome (MERS) dan Severe Acute Respiratory Syndrome (SARS) [2]. Virus corona ini telah memakan begitu banyak korban jiwa hampir di seluruh dunia, salah satunya adalah Indonesia. Pada tanggal 26 november 2021 data kasus terkonfirmasi covid telah mencapai 453.000 kasus. Hal ini telah menurun jika dibandingkan dengan jumlah kasus terkonfirmasi pada bulan juli yang mancapai 56.757.000 kasus [4]. Di wilayah NTT tepatnya di kabupaten Sikka tidak terlepas dari banyaknya jumlah kasus covid. Ada 21 kecamatan di wilayah kabupaten Sikka yang memiliki data kasus covid terkonfirmasi sembuh dan meninggal. Data yang didapat bervariasi ada kecamatan yang memiliki data sembuh yang banyak namun jumlah meninggal sedikit, tetapi ada pula yang memiliki data kesembuhan yang kurang namun jumlah meninggal meningkat, ada juga yang data sembuh banyak dan meninggal banyak. Dari data tersebut maka peneliti ingin mengelompokkan kecamatan mana saja yang paling banyak terdampak kasus covid 19. Adapun algoritma yang digunakan pada penelitian ini adalah algoritma K-Means yang akan membagi data menjadi cluster, diantaranya cluster pertama adalah kecamatan yang paling banyak kasus covid, cluster kedua adalah kecamatan yang sedang untuk kasus covidnya sedangkan pada cluster ketiga adalah kecamatan yang kurang untuk penyebaran kasus covidnya. Dari hasil klasterisasi pada 21 kecamatan di kabupaten Sikka terdapat data sembuh berjumlah 4593 kasus, dan data meninggal 91 kasus. Penelitian ini bertujuan untuk menentukan pengelompokkan wilayah berdasarkan tingkat penyebaran Covid-19 di wilayah Kabupaten Sikka menggunakan metode clustering algoritma K-Means.

3. METODE PENELITIAN

Penelitian dilakukan dengan menggunakan metode klasifikasi K-Means. Metode klasifikasi K-Means merupakan algoritma yang melakukan pengelompokan data berdasarkan data yang terdekat dengan titik pusat cluster [4]. Data yang digunakan dalam penelitian ini adalah data penyebaran virus Covid-19 di kabupaten Sikka yang diperoleh dari Dinas Kabupaten sikka yang direkap pada tanggal 28 November 2021. Data mining disebut juga Knowledge Discovery in Database (KDD) melakukan ekstraksi informasi dari tumpukan data. Proses pencarian informasi tersebut akan menemukan pola kecenderungan dari data yang kemudian hasil dari penambangan (mining) dapat menjadi informasi yang mudah dipahami. Analisis cluster merupakan salah satu teknik data mining yang bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok objek lainnya, sehingga objek yang berada dalam kelompok yang sama relatif lebih homogen daripada objek yang berada pada kelompok yang berbeda [5]. *K-Means* yaitu salah satu dari metode pengelompokan data nonhierarki (sekatan) yang dapat mempartisi data kedalam bentuk dua kelompok ataupun lebih [9]. *K-Means clustering* merupakan sebuah konstanta dari sejumlah *cluster* yang diinginkan, sedangkan Means atau dapat didefinisikan sebagai *cluster* adalah suatu nilai rata-rata dari sekumpulan populasi data [12]. Metode *K-Means* digunakan untuk mengelompokkan data ke dalam beberapa kelompok, data dalam satu kelompok mempunyai karakteristik yang sama dan mempunyai karakteristik yang berbeda dengan data dari kelompok yang lain [13]. Metode ini berusaha meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi data yang ada di *cluster* lainnya.

Algoritma K-Means adalah seperti berikut [6];

1. Tentukan k (jumlah cluster) yang ingin dibentuk
2. Bangkitkan k centroid (titik pusat cluster) awal secara random
3. Untuk setiap record , temukan pusat cluster terdekat
4. Untuk setiap k cluster , temukan pusat cluster ,dan update lokasi dari setiap pusat cluster dengan nilai centroid yang baru. Pusat cluster diperoleh dengan cara menghitung nilai rata-rata dari data-data yang berada pada cluster yang sama.
5. Kembali ke langkah 3 – 5 sampai konvergen.

Metode clustering menggunakan Algoritma K-means, ukuran kedekatan data dihitung menggunakan jarak Euclidean. Algoritma K-means bertujuan untuk meminimumkan jarak total Euclidean diantara setiap titik X_i , dan cluster terdekat yakni c_j [7]. Jarak Euclidean ditentukan dengan menggunakan persamaan 1 berikut ini [8].

$$d_{ij} = \sqrt{\sum_{k=0}^n \{x_{ik} - x_{jk}\}^2} \quad (1)$$

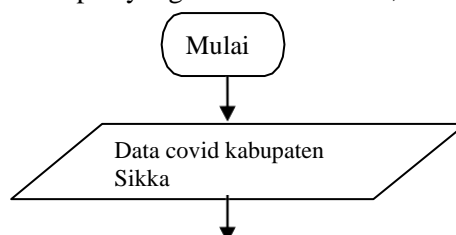
Keterangan: d_{ij} = Jarak antara data ke-i dan data ke-j n = Dimensi data x_{ik} = Koordinat data ke-i pada dimensi k x_{jk} = Koordinat data ke-j pada dimensi k
EuclideanDistance Formula untuk menghitung jarak antar dengan EuclideanDistance untuk dua titik dalam satu, dua, dan tiga dimensi secara berurutan ditunjukkan pada persamaan (2), (3), dan (4) sebagai berikut [9].

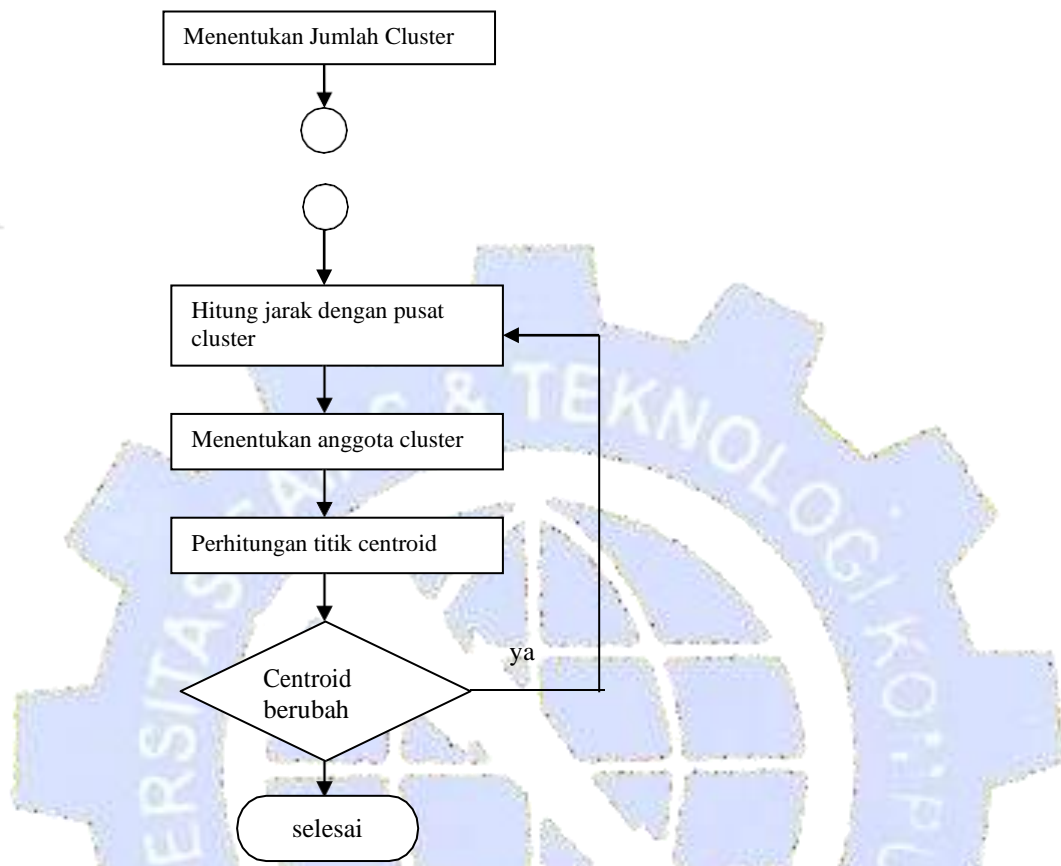
$$\sqrt{(x - y)^2} = |x - y| \quad (2)$$

$$d(p, q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2} \quad (3)$$

4. HASIL DAN PEMBAHASAN

Pada penelitian ini tahapan yang dilakukan adalah;





Gambar 1. Langkah-Langkah Pengklusteran Menggunakan Metode K-Means

Tabel 1 DATA KASUS KONFIRMASI PERKECAMATAN

Kecamatan	Sembuh	Meninggal
ALOK	1043	23
ALOK TIMUR	1059	20
ALOK BARAT	348	5
MAGEPANDA	73	2
PALUE	20	1
TALIBURA	114	2
WAIGETE	180	5
WAIBLAMA	36	1
MAPITARA	11	0
DORENG	63	3
BOLA	78	3
HEWOKLOANG	64	1
KEWAPANTE	207	6
KANGAE	281	3
NELLE	123	0
KOTING	116	3
NITA	309	7
LELA	202	2
MEGO	153	2
TANAWAWO	10	0

PAGA	103	2
TOTAL	4593	91

Dari data penyebaran Covid-19 21 Kecamatan di Kabupaten Sikka tersebut ditentukan tiga cluster sebagai titik centroid awal yaitu data ke-6 (Talibura), 12(Hewokloang), dan 17(Nita) dapat dilihat pada Tabel 1. *Centroid* awal yang telah ditentukan dapat dilihat pada Tabel 2. Setelah itu langkah selanjutnya adalah menghitung jarak antara *centroid* awal dengan titik *centroid*. Hasil yang didapatkan untuk menentukan jarak antara *centroid* awal dengan titik *centroid* menggunakan rumus *euclidian distance* yang hasilnya terdapat pada Tabel 3 ini, kemudian dipilih *centroid* mana dari masing-masing kode yang memiliki jarak terpendek.

Tabel 2. Centroid Awal

<i>centroid</i> ke 1	TALIBURA	114	2
<i>centroid</i> ke 2	HEWOKLOANG	64	1
<i>centroid</i> ke 3	NITA	309	7

Tabel 3. Hasil perhitungan cluster dengan titik centroid

no	c1	c2	c3	min	cluster
1	929.2373	979.2472	734.1744	734.1744	3
2	945.1714	995.1814	750.1127	750.1127	3
3	234.0192	284.0282	39.05125	39.05125	3
4	41	9.055385	236.053	9.055385	2
5	94.00532	44	289.0623	44	2
6	0	50.01	195.0641	0	1
7	66.06815	116.0689	129.0155	66.06815	1
8	78.00641	28	273.0659	28	2
9	103.0194	53.00943	298.0822	53.00943	2
10	51.0098	2.236068	246.0325	2.236068	2
11	36.01389	14.14214	231.0346	14.14214	2
12	50.01	0	245.0735	0	2
13	93.08598	143.0874	102.0049	93.08598	1
14	167.003	217.0092	28.28427	28.28427	3
15	9.219544	59.00847	186.1317	9.219544	1
16	2.236068	52.03845	193.0414	2.236068	1
17	195.0641	245.0735	0	0	3
18	88	138.0036	107.1168	88	1
19	39	89.00562	156.0801	39	1
20	104.0192	54.00926	299.0819	54.00926	2
21	11	39.01282	206.0607	11	1

Tabel 4. Keanggotaan Cluster Hasil Perhitungan Centroid pada Iterasi 1

	c1	c2	c3
			1
			1
			1
		1	
		1	
1			
1			
		1	
		1	
		1	
		1	
		1	
1			
			1

1		
1		
		1
1		
1		
	1	
1		

Tabel 5. Nilai Centroid dari Hasil Perhitungan Iterasi 1

centroid 1	149.75	2.75
centroid 2	44.375	1.375
centroid 3	11.6	11.6

Tabel 6. Nilai Centroid dari Hasil Perhitungan Iterasi 2

no	c1	c2	c3	min	cluster	ket
1	893.4795	998.8591	1031.463	893.4795	1	Berubah
2	909.4136	1014.796	1047.434	909.4136	1	Berubah
3	198.2628	303.6466	336.4647	198.2628	1	Berubah
4	76.75366	28.63182	62.14596	28.63182	2	Aman
5	129.7618	24.37788	13.52479	13.52479	3	Berubah
6	35.75787	69.62781	102.849	35.75787	1	Aman
7	30.33356	135.6734	168.5293	30.33356	1	Aman
8	113.7635	8.383391	26.60301	8.383391	2	Aman
9	138.7772	33.40331	11.61551	11.61551	3	Berubah
10	86.75036	18.69575	52.11449	18.69575	2	Aman
11	71.75044	33.66424	66.95461	33.66424	2	Aman
12	85.76786	19.62858	53.46139	19.62858	2	Aman
13	57.34217	162.6908	195.4802	57.34217	1	Aman
14	131.2502	236.6306	269.5372	131.2502	1	Berubah
15	26.89098	78.63702	112.0023	26.89098	1	Aman
16	33.75093	71.64343	104.7536	33.75093	1	Aman
17	159.3067	264.6848	297.4356	159.3067	1	Berubah
18	52.25538	157.6262	190.6419	52.25538	1	Aman
19	3.335416	108.6268	141.7255	3.335416	1	Aman
20	139.7771	34.40249	11.70982	11.70982	3	Berubah
21	46.75602	58.62833	91.90277	46.75602	1	Aman

Dari hasil perhitungan centroid pada iterasi 1 yaitu pada tabel 4, maka langkah selanjutnya dengan kembali menghitung nilai rata-rata anggota *cluster* sehingga didapatkan nilai *centroid* iterasi 1 yang dilihat pada Tabel 5. Hasil yang didapatkan di iterasi 1 ternyata terdapat perubahan nilai *centroid*, maka diharuskan melakukan perhitungan kembali menggunakan rumus *euclidian distance* untuk menentukan jarak antara *centroid* dengan titik *centroid* yang hasilnya terdapat pada Tabel 6, kemudian dipilih kembali *centroid* mana dari masing-masing kode yang memiliki jarak terpendek, hasil perhitungan tersebut dapat dilihat pada Tabel 6. Dari Tabel 6 terlihat bahwa terdapat kolom keterangan yang berisi pernyataan jika cluster pada iterasi 1 sama dengan cluster pada iterasi 2 maka akan muncul pernyataan “Aman”, namun jika tidak maka muncul pernyataan “Berubah”. Perhitungan untuk mendapatkan semua keterangan “Aman” berhenti pada iterasi 9 yang terlihat pada table 7.

Tabel 7. Nilai Centroid dari Hasil Perhitungan Iterasi 9

no	c1	c2	c3	min	cluster	ket
1	8.13941	788.7131	969.0838	8.13941	1	Aman
2	8.13941	804.6461	985.0192	8.13941	1	Aman
3	703.1936	93.50059	273.868	93.50059	2	Aman
4	978.1944	181.5196	1.24273	1.24273	3	Aman
5	1031.204	234.5287	54.15652	54.15652	3	Aman
6	937.2029	140.5253	39.84883	39.84883	3	Aman



Gambar 2. Grafik kecamatan cluster 1



Gambar 3. Grafik kecamatan cluster 2



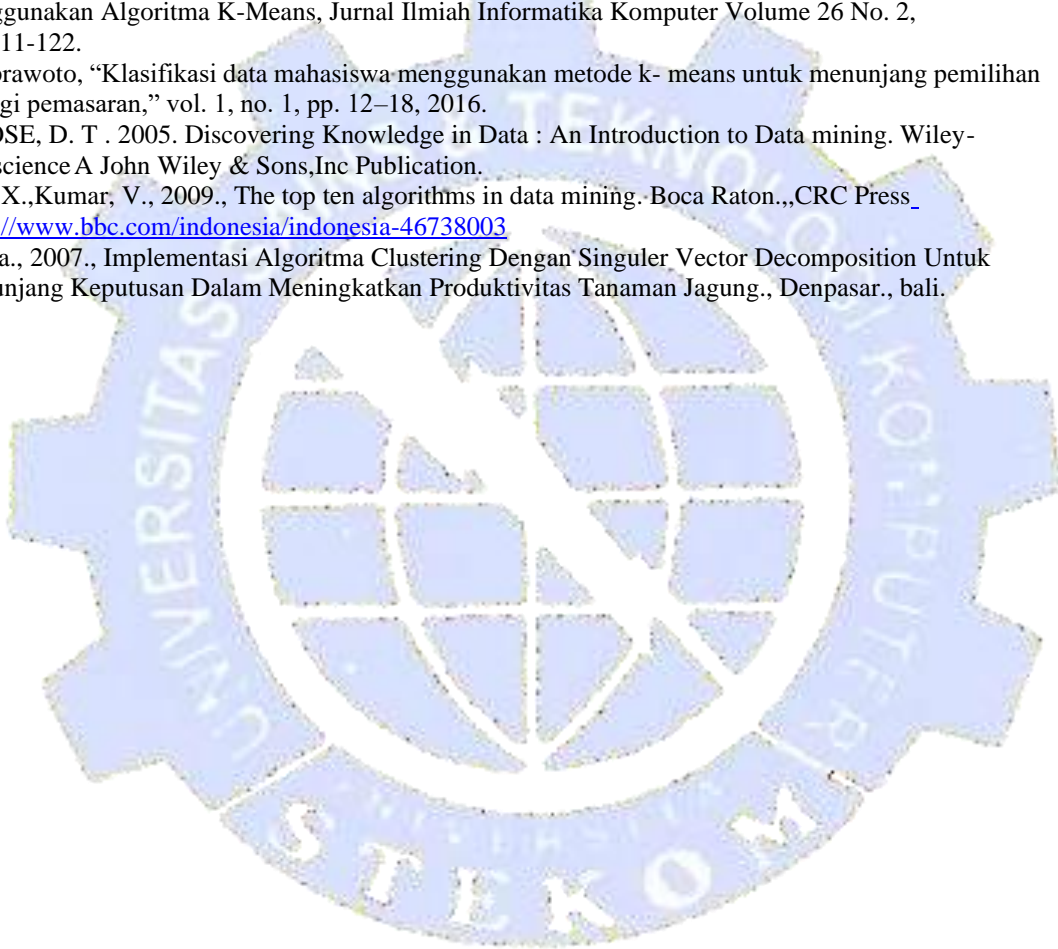
Gambar 2. Grafik kecamatan cluster 3

5. KESIMPULAN

Setelah dilakukan proses pengelompokan wilayah Kecamatan terdampak Covid-19 di Kabupaten Sikka dengan menggunakan metode K-Means dapat disimpulkan bahwa cluster dibagi menjadi 3 yaitu cluster 1 yang berisi data kecamatan yang berdampak covid tertinggi terdapat 2 kecamatan yaitu Alok dan Alok Timur. Cluster ke 2 yang berisi data kecamatan yang berdampak covid sedang terdapat 6 Kecamatan. Cluster 3 yang berdampak covid rendah terdapat 13 kecamatan. Pandemi Covid-19 masih berlangsung maka diperlukan pengalihan data lanjutan terhadap wilayah terdampak Covid-19 di Kabupaten Sikka dengan menggunakan data-data yang telah diperbaharui setiap harinya berdasarkan kasus yang terjadi di masing-masing kecamatan.

DAFTAR PUSTAKA

- [1] Wahidah,2020, Pandemi Covid-19: Analisis Perencanaan Pemerintah dan Masyarakat dalam Berbagai Upaya Pencegahan, Jurnal Manajemen dan Organisasi (JMO), Vol. 11 No. 3, Hal. 179-188
- [2] C. C. Lai, T. P. Shih, W. C. Ko, H. J. Tang, and P. R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," Int. J. Antimicrob. Agents, vol. 55, no. 3, p. 105924, 2020, doi: 10.1016/j.ijantimicag.2020.105924.
- [3] peta sebaran covid 19, <https://covid19.go.id/peta-sebaran-covid19>, diakses pada tanggal 29 November 2021
- [4] Alfianti,2021, Pengelompokan Wilayah Penyebaran Covid-19 Di Kabupaten Karawang Menggunakan Algoritma K-Means, Jurnal Ilmiah Informatika Komputer Volume 26 No. 2, Hal.111-122.
- [5] T. Suprawoto, "Klasifikasi data mahasiswa menggunakan metode k- means untuk menunjang pemilihan strategi pemasaran," vol. 1, no. 1, pp. 12–18, 2016.
- [6] LAROSE, D. T . 2005. Discovering Knowledge in Data : An Introduction to Data mining. Wiley- Interscience A John Wiley & Sons, Inc Publication.
- [7] Wu, X.,Kumar, V., 2009., The top ten algorithms in data mining. Boca Raton.,CRC Press_ <https://www.bbc.com/indonesia/indonesia-46738003>
- [8] Agusta., 2007., Implementasi Algoritma Clustering Dengan Singular Vector Decomposition Untuk Menunjang Keputusan Dalam Meningkatkan Produktivitas Tanaman Jagung., Denpasar., bali.



PUBLIKASI ILMIAH