

IMPLEMENTASI K-MEANS CLUSTERING ANALYSIS UNTUK MENGELOMPOKKAN KELURAHAN-KELURAHAN DI DKI JAKARTA BERDASARKAN JUMLAH POSITIF COVID-19.

Nurul Qomariasih

Politeknik Siber dan Sandi Negara (Poltek SSN) Bogor, Jawa Barat, Indonesia

Email: nurul.qomariasih@poltekssn.ac.id

INFO ARTIKEL	ABSTRAK
Diterima 21 Juni 2021 Direvisi 2 Juli 2021 Disetujui 15 Juli 2021	Covid 19 merupakan penyakit yang disebabkan oleh Virus Corona, menyerang sistem pernapasan, dan memperburuk kondisi seseorang dengan penyakit bawaan. Hingga saat ini, pemerintah masih menetapkan kasus wabah covid 19 sebagai kondisi darurat dan penetapan protokol kesehatan kepada warganya. DKI Jakarta merupakan salah satu daerah dengan kasus tertinggi covid 19. Oleh karena itu, penelitian ini melakukan pengelompokan atau <i>clustering</i> kelurahan-kelurahan yang ada di provinsi DKI Jakarta ke dalam kelompok kelurahan dengan kasus sangat tinggi, sedang, atau rendah menggunakan metode <i>K-Means Clustering</i> . Pengelompokan ini bertujuan sebagai pemberian informasi kepada pemerintah dan sebagai salah satu bahan pertimbangan dalam menangani kasus covid 19 serta prioritas penyediaan sarana dan prasarana medis. Penelitian <i>K-Means Clustering</i> dengan data yang telah di standarisasi menggunakan jarak Euclidean berhasil mengelompokkan kelurahan-kelurahan di DKI Jakarta menjadi 3 Kelompok.
Kata Kunci: Covid 19; Pengelompokan; K-Means	ABSTRACT <i>Covid 19 is a disease caused by the Corona Virus, attacking the respiratory system, and worsening the condition of a person with congenital disease. Until now, the government has still determined the case of the COVID-19 outbreak as an emergency and established health protocols for its citizens. DKI Jakarta is one of the areas with the highest cases of covid 19. Therefore, this study is grouping or clustering urban villages in DKI Jakarta province into groups of urban villages with very high, medium, or low cases using the K-Means Clustering method . This grouping aims to provide information to the government and as one of the considerations in dealing with COVID-19 cases as well as prioritizing the provision of medical facilities and infrastructure. The K-Means Clustering study with standardized data using Euclidean distance succeeded in grouping urban villages in DKI Jakarta into 3 groups.</i>
Keywords: Covid 19; Clustering; K-Means	

Pendahuluan

K-Means adalah salah satu teknik dalam *Data Mining* yang bertujuan untuk mengelompokkan data yang mempunyai kemiripan karakteristik antara data yang satu

dengan data lainnya (Wardhani & Khrisna, 2016). Teknik ini bertujuan untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, meminimalkan variasi suatu kelompok dan

How to cite:	Nurul Qomariasih (2021) Implementasi K-Means Clustering Analysis untuk mengelompokkan kelurahan di DKI Jakarta berdasarkan jumlah positif covid-19. <i>Jurnal Syntax Transformation</i> 2(7). https://doi.org/10.46799/jst.v2i7.336
E-ISSN:	2721-2769
Published by:	Ridwan Institute

memaksimalkan variasi antar kelompok (Maulida, 2018). Sedangkan *Data Mining* adalah suatu proses pencarian pola, hubungan, penggalian nilai tambah dari data yang berukuran besar berupa pengetahuan dengan tujuan menemukan hubungan dan menyederhanakan data agar didapatkan informasi yang mudah dipahami serta bermanfaat (Abdurrahman, 2016). Penelitian ini juga menggunakan data besar atau data dengan jumlah observasi besar, sehingga disebut dengan *Big Data*. *Big Data* merupakan data yang melebihi proses kapasitas dari konvensi sistem database yang ada (E. Dumbill, 2012). Teknik / algoritma *K-Means* bekerja dengan cara membagi data dalam sejumlah *cluster* untuk dianalisis faktor kesamaan (*similarity*) maupun ketidaksamaannya (*dissimilarity*) yang melekat pada kumpulan data tersebut, dan selanjutnya dianalisis pola keterhubungan antar data (Larose & Larose, 2014).

Covid 19 dari awal kemunculannya hingga saat ini masih dianggap sebagai pandemi, karena sangat banyak yang menderita dan yang meninggal karena penyakit covid 19. Serta tidak ada satu negarapun yang tidak pernah terjangkit virus *Corona* penyebab covid 19 ini. Jumlah penduduk yang dikonformasi statusnya, seperti ODP, PDP, serta positif berbeda dari satu negara dengan negara lainnya, dikarenakan perbedaan kebijakan dan kemampuan kepala negaranya dalam menangani tindakan pencegahan dan penindakan terhadap pandemi.

Penelitian ini bertujuan untuk mendapatkan dan mengolah data suspek dan positif covid 19 di kelurahan di DKI Jakarta menggunakan *Data Mining* yaitu *K-Means Clustering*. Data akan dikelompokkan berdasarkan kesamaan yang terdekat. Kelurahan yang jumlah kasusnya tinggi akan berkelompok dengan kelurahan yang mirip, yaitu yang jumlah kasusnya tinggi juga. Sehingga pemerintah atau organisasi lain

dapat mengetahui wilayah prioritas bantuan Covid 19 dan mengurangi dampak pandemi ini sebanyak mungkin (Indraputra & Fitriana, 2020).

Metode Penelitian

A. Data Penelitian

Data yang digunakan dalam penelitian ini berupa data kuantitatif yaitu jumlah suspek dan jumlah positif Covid-19 di 267 kelurahan di DKI Jakarta pada tanggal 8 September 2020. Data yang digunakan dalam penelitian berasal dari halaman *website Open Data Jakarta* (<https://data.jakarta.go.id/dataset/reakap-data-covid-19-per-kelurahan-provinsi-dki-jakarta-bulan-september-2020>). Data set yang didapat adalah riwayat Covid-19 per kelurahan di Provinsi DKI Jakarta pada tanggal 8 September 2020. Dataset tersebut berisi berbagai macam variabel dan merupakan penjumlahan dari variabel yang ada seperti : ODP (Proses Pemantauan + Selesai Pemantauan), PDP (Masih Dirawat + Pulang dan Sehat), Positif (Dirawat + Sembuh + Meninggal + *Self Isolation*).

Penelitian ini dilakukan dengan beberapa langkah: pertama, dilakukan *preprocessing data* yang meliputi pengumpulan data dari basis data, pembersihan data, serta transformasi data (Fayyad et al., 1996). Data ditransformasi ke dalam bentuk *Z score* atau biasa disebut normalisasi. Data hasil normalisasi akan berada pada rentang 0 sampai 1. Proses normalisasi ini sangat penting untuk memastikan bahwa tidak ada parameter yang mendominasi dalam perhitungan jarak antar data (Hann & Kamber, 2000). Tahap selanjutnya adalah uji asumsi multikolinearitas dengan melakukan uji pearson untuk memastikan variabel yang digunakan tidak saling berhubungan/saling bebas.

B. Penelitian Terkait

Berikut merupakan beberapa penelitian menggunakan metode *k-means clustering* untuk mengelompokkan data:

Tahun	Judul	Subjek Observasi
2020	Analisis <i>K-Means Clustering</i> pada Data Sepeda Motor	Sepeda Motor merek Honda dan Yamaha
2020	<i>K-Means Clustering</i> Data Covid-19	Data kasus covid-19 dari seluruh provinsi di dunia
2021	<i>Clustering</i> Kualitas Kinerja Karyawan pada Perusahaan Bahan Kimia Menggunakan Algoritma <i>K-Means</i>	Pegawai PT Clariant Adsorbent Indonesia

C. *K-Means Clustering*

K-Means Clustering merupakan teknik pengelompokkan data non-hirarki dan bersifat *unsupervised* (tanpa arahan), yaitu tidak mempunyai anggota *cluster* yang pasti. Namun, terkadang peneliti menentukan jumlah *cluster* berdasarkan pengalaman atau melalui metode tertentu, sehingga *K-Means Clustering* disebut juga sebagai *semi-supervised classification* (Agusta, 2007). Langkah pertama adalah menentukan *k* sebagai jumlah clusternya. Selanjutnya mengukur jarak kesamaan menggunakan jarak Euclidean. Sarwono dalam bukunya bahwa algoritma *K-Means* adalah sebagai berikut (Rohmawati et al., 2015):

1. Menentukan nilai *k* sebagai jumlah *cluster* yang ingin dibentuk
2. Menentukan ilai acak untuk pusat *cluster* awal centroid sebanyak *k* untuk menghitung jarak setiap data input terhadap masing-masing

centroid dengan menggunakan rumus jarak *Euclid*, yaitu:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

x_i = data kriteria

μ_j = *centroid* pada *cluster* ke – j

3. Mengelompokkan setiap data berdasarkan kedekatannya dengan *centroid* atau mencari jarak terkecil
4. Memperbaharui nilai *centroid* baru, nilai diperoleh dari rata-rata *cluster* yang bersangkutan.

$$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum_{j \in S_j} x_j$$

$\mu_j(t+1)$ = *centroid* baru pada iterasi (t + 1)

N_{S_j} = Data pada *cluster* S_j

5. Apabila data setiap *cluster* belum berhenti, lakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

Hasil Dan Pembahasan

Berdasarkan grafik jumlah kasus positif di DKI Jakarta tahun 2020 bulan September hingga Oktober, terjadi pelonjakan jumlah yang signifikan pada tanggal 8 September (Qomariasih, 2021). Oleh karena itu, kejadian lonjakan tersebut akan dianalisis dan dihasilkan tingkat keparahan di masing-masing kelurahan, yaitu Tinggi (Parah), Sedang, Rendah.

Berikut ditampilkan hasil pengujian asumsi multikolinearitas bagi kedua variabel yang digunakan (Tabel 1):

Tabel 1
Nilai VIF Variabel Suspek dan Positif Covid 19

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	21,23	5,19	4,09	0,000	
Suspek	0,3445	0,0183	18,78	0,000	1,00

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	68,0	11,0	6,20	0,000	
Positif	1,6574	0,0882	18,78	0,000	1,00

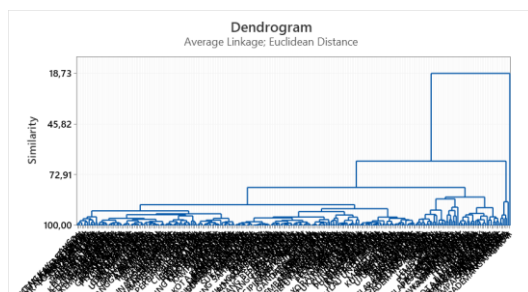
Jika hipotesis bagi uji multikolinearitas ini adalah:

H0: Tidak terjadi multikolinearitas

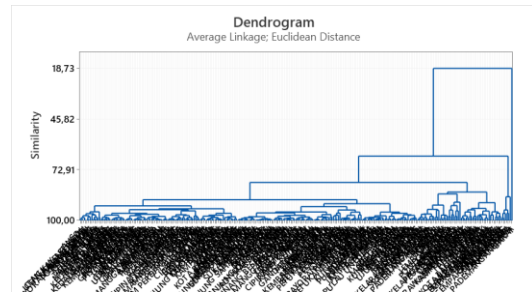
H1: Terjadi multikolinearitas

Dengan kriteria uji tolak H0 jika VIF >5, maka berdasarkan tabel 1, kesimpulannya terima H0 untuk variabel suspek maupun variabel positif.

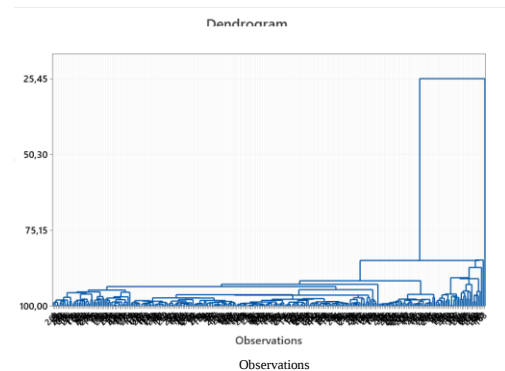
Penentuan jumlah *cluster* dengan *K-Means* dapat ditentukan secara subjektif oleh peneliti. Namun pada penelitian ini tetap dibuat dendrogram dari variabel jumlah suspek dan variabel jumlah positif covid 19 tiap kelurahan yang telah distandarisasi menjadi nilai Z. Dendrogram dibuat dengan berbagai *Linkage Method* sebagai referensi penentuan jumlah *cluster* pada analisis *K-Means* (Grafik 1,2, dan 3). Pada Tabel 1, dendrogram dengan *Average Linkage* memiliki garis yang lebih panjang dari garis lainnya yaitu ada sebanyak 3 garis pertama, sehingga kalau diambil *k* menjadi 4 *cluster*.



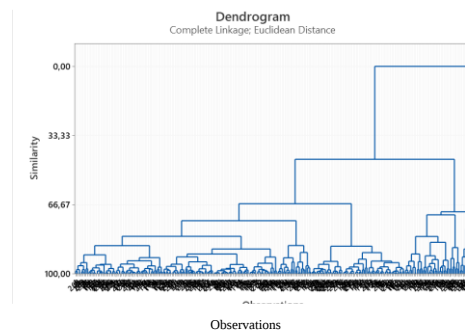
Grafik 1
Dendrogram Kelurahan dengan Jarak Euclid dan Average Linkage



Grafik 2
Dendrogram Kelurahan dengan Jarak Euclid dan Centroid Linkage



Gambar 3
Dendrogram Kelurahan dengan Jarak Euclid dan Complete Linkage



Penentuan jumlah *cluster* yang akan diambil dapat dilihat dari panjang garis dendrogram. Garis yang dekat menandakan *similarity* yang paling dekat, semakin jauh dan semakin panjang garisnya, maka *similarity* nya atau kesamaannya juga semakin jauh. Berdasarkan tiga grafik dendrogram di atas, maka kita bisa mengambil *cluster* sebanyak 2, 3, atau 4. Agar sesuai dan relevan dengan konteks penelitian, serta mengingat ragam yang besar dari data yang

ada, maka akan diambil jumlah *cluster* sebanyak 3, dengan penamaan *cluster* nya yaitu parah, sedang dan rendah. Berikut merupakan hasil pengolahan menggunakan *K-Means Clustering*:

Tabel 2
Informasi Jarak Cluster dengan titik pusat

	Number of observations	sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	34	81,952	1,171	6,623
Cluster2	120	63,070	0,647	1,783
Cluster3	113	34,409	0,494	1,111

Tabel 3
Informasi Pusat Cluster

Variable	Cluster1	Cluster2	Cluster3	Grand centroid
Positif	1,9343	0,1514	-0,7428	0,0000
suspek	1,6319	0,2398	-0,7456	0,0000

Terlihat dari Tabel 2 di atas bahwa terdapat 34 kelurahan berada pada *Cluster* 1, 120 kelurahan berada pada *Cluster* 2, dan 113 kelurahan berada pada *Cluster* 3. Kolom ketiga (*Within cluster sum of square*) memberikan informasi jarak kuadrat antar *Cluster*. Kolom keempat (*Average distance from centroid*) adalah jarak rata-rata tiap kelurahan pada satu *cluster* ke pusatnya (*centroid*). Sedangkan kolom terakhir (*Maximum distance from centroid*) menunjukkan jarak terjauh kelurahan yang berada pada masing-masing *cluster* ke pusat *clusternya*. Titik pusat (*centroid*) ditampilkan pada Tabel 3. Adapun hasil pengelompokkan kelurahan-kelurahan di DKI Jakarta ditunjukkan melalui tabel berikut:

No.	Kelurahan	Cluster
1	ANCOL	2
2	ANGKE	2
3	BALE KAMBANG	3

No.	Kelurahan	Cluster
4	BALI MESTER	3
5	BAMBU APUS	3
6	BANGKA	3
7	BARU	3
8	BATU AMPAR	2
9	BENDUNGAN HILIR	2
10	BIDARA CINA	2
11	BINTARO	2
12	BUKIT DURI	3
13	BUNGUR	3
14	CAKUNG BARAT	3
15	CAKUNG TIMUR	3
16	CAWANG	3
17	CEGER	3
18	CEMPAKA BARU	1
19	CEMPAKA PUTIH BARAT	1
20	CEMPAKA PUTIH TIMUR	1
21	CENGKARENG BARAT	2
22	CENGKARENG TIMUR	1
23	CIBUBUR	2
24	CIDENG	2
25	CIGANJUR	2
26	CIJANTUNG	2
27	CIKINI	3
28	CIKOKO	3
29	CILANDAK BARAT	2
30	CILANDAK TIMUR	2
31	CILANGKAP	3
32	CILILITAN	2
33	CILINCING	1
34	CIPAYUNG	2
35	CIPEDAK	2
36	CIPETE SELATAN	3
37	CIPETE UTARA	2
38	CIPINANG	2
39	CIPINANG BESAR SELATAN	3
40	CIPINANG BESAR UTARA	3
41	CIPINANG	3
42	CEMPEDAK	2
43	CIPINANG MELAYU	2
44	CIPINANG MUARA	2
45	CIPULIR	2
46	CIRACAS	2
47	DUKUH	3
48	DUREN SAWIT	1
	DUREN TIGA	3

No.	Kelurahan	Cluster	No.	Kelurahan	Cluster
49	DURI KEPA	2	96	KARANG ANYAR	2
50	DURI KOSAMBI	2	97	KARET	3
51	DURI PULO	3	98	KARET KUNINGAN	3
52	DURI SELATAN	3	99	KARET SEMANGGI	3
53	DURI UTARA	3	100	KARET TENGSIN	3
54	GALUR	3	101	KARTINI	2
55	GAMBIR	3	102	KAYU MANIS	3
	GANDARIA		103	KAYU PUTIH	2
56	SELATAN	3	104	KEAGUNGAN	3
57	GANDARIA UTARA	2	105	KEBAGUSAN	2
58	GEDONG	2		KEBAYORAN LAMA	
59	GELORA	3	106	SELATAN	2
60	GLODOK	3		KEBAYORAN LAMA	
61	GONDANGDIA	3	107	UTARA	3
62	GROGOL	2	108	KEBON BARU	3
63	GROGOL SELATAN	2	109	KEBON BAWANG	1
64	GROGOL UTARA	2	110	KEBON JERUK	1
65	GUNTUR	3	111	KEBON KACANG	1
66	GUNUNG	3	112	KEBON KELAPA	3
	GUNUNG SAHARI		113	KEBON KOSONG	2
67	SELATAN	3	114	KEBON MANGGIS	3
	GUNUNG SAHARI		115	KEBON MELATI	1
68	UTARA	3	116	KEBON PALA	3
	HALIM PERDANA		117	KEBON SIRIH	3
69	KUSUMAH	2		KEDAUNG KALI	
70	HARAPAN MULIA	2	118	ANGKE	2
71	JAGAKARSA	1	119	KEDOYA SELATAN	3
72	JATI	2	120	KEDOYA UTARA	3
73	JATI PADANG	3	121	KELAPA DUA	3
74	JATI PULO	2		KELAPA DUA	
75	JATINEGARA	2	122	WETAN	2
76	JATINEGARA KAUM	3		KELAPA GADING	
77	JELAMBAR	2	123	BARAT	1
78	JELAMBAR BARU	3		KELAPA GADING	
79	JEMBATAN BESI	2	124	TIMUR	2
80	JEMBATAN LIMA	3	125	KEMANGGISAN	2
81	JOGLO	3	126	KEMAYORAN	2
82	JOHAR BARU	1		KEMBANGAN	
83	KALI ANYAR	2	127	SELATAN	3
84	KALI BARU	2		KEMBANGAN	
85	KALIBATA	2	128	UTARA	2
86	KALIDERES	1	129	KENARI	2
87	KALISARI	2	130	KERENDANG	3
88	KAMAL	3	131	KLENDER	1
89	KAMAL MUARA	3	132	KOJA	2
90	KAMPUNG BALI	3		KOTA BAMBU	
	KAMPUNG		133	SELATAN	2
91	MELAYU	3		KOTA BAMBU	
92	KAMPUNG RAWA	2	134	UTARA	2
93	KAMPUNG TENGAH	2	135	KRAMAT	2
94	KAPUK	1	136	KRAMAT JATI	2
95	KAPUK MUARA	3	137	KRAMAT PELA	3

Implementasi *K-Means Clustering Analysis* Untuk Mengelompokkan Kelurahan-Kelurahan Di
DKI Jakarta Berdasarkan Jumlah Positif Covid-19.

No.	Kelurahan	Cluster	No.	Kelurahan	Cluster
138	KRUKUT	3	185	PESANGGRAHAN	2
139	KUNINGAN BARAT	3	186	PETAMBURAN	1
140	KUNINGAN TIMUR	3	187	PETOGOGAN	3
141	KWITANG	3	188	PETOJO SELATAN	2
142	LAGOA	1	189	PETOJO UTARA	3
143	LEBAK BULUS	2		PETUKANGAN	
144	LENTENG AGUNG	2	190	SELATAN	3
145	LUBANG BUAYA	2		PETUKANGAN	
146	MAKASAR	3	191	UTARA	2
147	MALAKA JAYA	2	192	PINANG RANTI	3
148	MALAKA SARI	3	193	PINANGSIA	3
	MAMPANG		194	PISANGAN BARU	2
149	PRAPATAN	2	195	PISANGAN TIMUR	2
150	MANGGA BESAR	3	196	PLUIT	2
	MANGGA DUA		197	PONDOK BAMBU	1
151	SELATAN	2	198	PONDOK KELAPA	1
152	MANGGARAI	3	199	PONDOK KOPI	2
	MANGGARAI		200	PONDOK LABU	3
153	SELATAN	3	201	PONDOK PINANG	2
154	MAPHAR	2	202	PONDOK RANGGON	3
155	MARUNDA	3	203	PULAU HARAPAN	3
156	MELAWAI	3	204	PULAU KELAPA	3
157	MENTENG	2	205	PULAU PANGGANG	3
158	MENTENG ATAS	2	206	PULAU PARI	3
159	MENTENG DALAM	2	207	PULAU TIDUNG	3
160	MERUYA SELATAN	3		PULAU UNTUNG	
161	MERUYA UTARA	2	208	JAWA	3
162	MUNJUL	3	209	PULO	3
	PADEMANGAN		210	PULO GADUNG	2
163	BARAT	1	211	PULO GEBANG	2
	PADEMANGAN		212	RAGUNAN	2
164	TIMUR	2	213	RAMBUTAN	2
165	PAL MERIAM	2		RAWA BADA	
166	PALMERAH	1	214	SELATAN	1
167	PANCORAN	2		RAWA BADA	
168	PAPANGGO	2	215	UTARA	2
169	PASAR BARU	3	216	RAWA BARAT	3
170	PASAR MANGGIS	3	217	RAWA BUAYA	2
171	PASAR MINGGU	2	218	RAWA BUNGA	3
172	PASEBAN	2	219	RAWA JATI	3
173	PEGADUNGAN	2	220	RAWA TERATE	3
174	PEGANGSAAN	1	221	RAWAMANGUN	2
175	PEGANGSAAN DUA	1	222	RAWASARI	1
176	PEJAGALAN	2	223	ROA MALAKA	3
177	PEJATEN BARAT	2	224	ROROTAN	3
178	PEJATEN TIMUR	2	225	SELONG	3
179	PEKAYON	2	226	SEMANAN	2
180	PEKOJAN	3	227	SEMPER BARAT	1
181	PELA MAMPANG	2	228	SEMPER TIMUR	2
182	PENGADEGAN	3	229	SENAYAN	3
183	PENGGILINGAN	1	230	SENEN	3
184	PENJARINGAN	1	231	SERDANG	2

No.	Kelurahan	Cluster	Kesimpulan
232	SETIA BUDI	3	Metode <i>K-Means Clustering</i> mampu mengelompokkan kelurahan-kelurahan di DKI Jakarta berdasarkan tingkat keparahan Covid 19. Data diperoleh dari publikasi pemerintah DKI Jakarta melalui <i>website Open Data Jakarta</i> . Dari penelitian ini, diharapkan pemerintah dan khalayak dapat tetap waspada khususnya jika berada atau tinggal di kelurahan dengan tingkat keparahan yang tinggi.
233	SETU	3	
234	SLIPI	2	
235	SRENGSENG	2	
236	SAWAH	2	
	SUKABUMI		
237	SELATAN	3	
238	SUKABUMI UTARA	3	
239	SUKAPURA	1	
240	SUMUR BATU	2	
241	SUNGAI BAMBU	3	
242	SUNTER AGUNG	1	
243	SUNTER JAYA	1	
244	SUSUKAN	3	
245	TAMAN SARI	3	
246	TAMBORA	3	
247	TANAH SEREAL	2	
248	TANAH TINGGI	1	
249	TANGKI	3	
250	TANJUNG BARAT	2	
	TANJUNG DUREN		
251	SELATAN	2	
	TANJUNG DUREN		
252	UTARA	3	
253	TANJUNG PRIOK	2	
254	TEBET BARAT	2	
255	TEBET TIMUR	3	
256	TEGAL ALUR	2	
257	TEGAL PARANG	2	
258	TOMANG	2	
259	TUGU SELATAN	2	
260	TUGU UTARA	1	
261	UJUNG MENTENG	3	
262	ULUJAMI	2	
	UTAN KAYU		
263	SELATAN	2	
	UTAN KAYU		
264	UTARA	2	
265	UTAN PANJANG	2	
266	WARAKAS	2	
267	WIJAYA KUSUMA	2	

Berdasarkan tabel di atas, kelurahan yang berada pada *cluster* 1 merupakan kelurahan dengan tingkat kasus covid 19 yang paling tinggi, diikuti *cluster* 2 dengan kondisi sedang, dan *cluster* 3 rendah.

BIBLIOGRAFI

- Abdurrahman, G. (2016). Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 1(2). [Google Scholar](#)
- Agusta, Y. (2007). K-means-penerapan, permasalahan dan metode terkait. *Jurnal Sistem Dan Informatika*, 3(1), 47–60. [Google Scholar](#)
- E. Dumbill. (2012). *Big Data Now Current Perspective*. O'Reilly Media.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37. [Google Scholar](#)
- Hann, J., & Kamber, M. (2000). *Data mining: concepts and techniques*. Morgan Kaufman Publish. [Google Scholar](#)
- Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275–282. [Google Scholar](#)
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons. [Google Scholar](#)
- Maulida, L. (2018). Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan

- di Prov. DKI Jakarta dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167–174. [Google Scholar](#)
- Qomariasih, N. (2021). Peramalan Kasus Covid-19 di DKI Jakarta dengan Model Arima. *Jurnal Syntax Transformation*, 2(6), 849–855. [Google Scholar](#)
- Rohmawati, N., Defiyanti, S., & Jajuli, M. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 1(2). [Google Scholar](#)
- Wardhani, A. K., & Khrisna, A. (2016). Implementasi Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Kajen Pekalongan. *J. Transform*, 14(1), 30–37. [Google Scholar](#)

Copyright holder :

Nurul Qomariasih (2021)

First publication right :

Journal Syntax Transformation

This article is licensed under:

