

Clustering Kabupaten/Kota di Jawa Tengah Tahun 2022 berdasarkan Jumlah Kasus Kemunculan Penyakit dengan Algoritma K-Means

Nadiya Azhar Mufid¹, Derra Risqi Nurulita Hanum², Aldy Hasbiya Sidiq³

¹ Universitas Negeri Yogyakarta; nadiyaazhar.2020@student.uny.ac.id

² Universitas Negeri Yogyakarta; derrarisqi.2020@student.uny.ac.id

³ Universitas Negeri Yogyakarta; aldyhasbiya.2020@student.uny.ac.id

Abstrak: Penelitian ini bertujuan untuk melakukan *clustering* atau pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit-penyakit tertentu pada tahun 2022 menggunakan algoritma K-Means. Hasil penelitian diperoleh 3 *cluster* yaitu tinggi, sedang, dan rendah untuk 29 Kabupaten dan 6 Kota. Persentase untuk *cluster* 1 sebesar 34,29% yang terdiri atas 10 kabupaten dan 2 kota, *cluster* 2 sebesar 40,00% yang terdiri atas 11 kabupaten dan 3 kota, dan *cluster* 3 sebesar 25,71% yang terdiri atas 8 kabupaten dan 1 kota. Hasil *clustering* tersebut dapat digunakan sebagai dasar untuk mengambil keputusan strategis yang efektif dalam pengembangan upaya pencegahan dan penanggulangan penyakit di setiap wilayah.

Keywords: analisis pengelompokan, kejadian penyakit, algoritma k-means

DOI:

<https://doi.org/10.47134/ppm.v1i1.107>

*Correspondence: Nadiya Azhar Mufid

Email:

nadiyaazhar.2020@student.uny.ac.id

Received: 18-09-2023

Accepted: 16-10-2023

Published: 20-11-2023

Abstract: This research aims to conduct clustering or grouping of Regencies/Cities in Central Java Province based on the number of occurrences of specific diseases in 2022 using the K-Means algorithm. The research results obtained 3 clusters, namely high, medium, and low for 29 Regencies and 6 Cities. The percentage for cluster 1 is 34.29%, consisting of 10 regencies and 2 cities, cluster 2 is 40.00%, consisting of 11 regencies and 3 cities, and cluster 3 is 25.71%, consisting of 8 regencies and 1 city. These clustering results can be used as a basis for making effective strategic decisions in the development of prevention and control efforts for diseases in each region.

Keywords: clustering analysis, disease occurrence, k-means algorithm



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Pendahuluan

Di tengah kemajuan informasi dan teknologi saat ini, pengolahan dan analisis data menjadi hal yang penting dan perlu dilakukan dalam berbagai bidang, termasuk salah satunya bidang kesehatan. Kesehatan merupakan salah satu bidang yang harus menjadi prioritas dan patut diperhatikan oleh pemerintah dalam upaya mensejahterakan rakyatnya (Efendi et al., 2019). Dalam hal ini, analisis data terkait kesehatan dapat memberikan kontribusi kepada pemerintah dan para ahli kesehatan dalam melihat dan memahami secara nyata pola dan tren kemunculan suatu penyakit tertentu di suatu wilayah atau daerah. Di dalam Undang-undang No. 36 Tahun 2009 tentang Kesehatan, menyatakan bahwa kesehatan merupakan hak asasi manusia yang menjadi hak fundamental bagi setiap warga negara dan mutlak untuk dipenuhi. Pembangunan pada bidang kesehatan merupakan salah satu upaya dalam memenuhi hak dasar setiap rakyat pada suatu negara dengan memperoleh pelayanan Kesehatan (Sitinjak et al., 2022).

Berdasarkan data yang dikeluarkan oleh Dinas Kesehatan Provinsi Jawa Tengah Tahun 2022, Provinsi Jawa Tengah menghadapi kondisi dan tantangan dalam mengatasi penyakit tertentu yang muncul pada berbagai wilayah di Kabupaten/Kota. Dalam upaya membantu mengidentifikasi pola dan karakteristik kemunculan penyakit tertentu pada masing-masing Kabupaten/Kota di Provinsi Jawa Tengah, akan dilakukan *clustering* menggunakan algoritma K-Means (Abualigah, 2018; Afzal, 2020; Gianniou, 2018; Sun, 2019; Yang, 2019; Zhao, 2018; Zhu, 2019). Metode K-Means *Clustering* merupakan salah satu metode algoritma *clustering* yang dilakukan dengan mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok (Alam, 2019; Borlea, 2021; Dubey, 2018; Ismkhan, 2018; Rezaee, 2021; Sinaga, 2020; Xie, 2019; Yu, 2018; Zhang, 2018).

Pada penelitian yang dilakukan oleh (Wibowo & Mulyastuti, 2022), penggunaan algoritma K-Means bertujuan dalam mengelompokkan fasilitas kesehatan yang ada di Pemerintahan Provinsi DKI Jakarta. Hasil penelitian diperoleh hasil berupa 3 *cluster*, yaitu tinggi, sedang, dan rendah di mana pada *cluster* rendah dengan fasilitas kesehatan paling sedikit berada di Kepulauan Seribu dengan jumlah 1 rumah sakit, 4 puskesmas kelurahan, 2 puskesmas kecamatan, dan 10 tempat tidur. Dari hasil penelitian yang diperoleh dapat disimpulkan bahwa masih kurangnya fasilitas kesehatan di Kepulauan Seribu, sehingga Pemerintah Provinsi DKI Jakarta dapat menambah fasilitas kesehatan di Kepulauan Seribu terlebih daerah tersebut termasuk daerah terpencil yang jauh dari daratan pusat kota untuk akses kesehatan.

Kemudian penelitian oleh (Sembiring, 2021) penggunaan algoritma K-Means (*clustering*) dalam pemetaan desa yang terjangkit Demam Berdarah Dengue. Pemetaan bertujuan untuk mencari daerah yang rawan terjangkit penyakit DBD dan daerah yang jarang terjangkit penyakit DBD. Hasil penelitian diperoleh 3 *cluster* untuk tingkat penyebaran penyakit DBD yaitu *cluster* tinggi, sedang, dan rendah di mana *cluster* tinggi dapat ditemukan di desa Kecamatan Setia Janji yang paling sering dan paling banyak

masyarakatnya terjangkau penyakit DBD yaitu desa 2 dan 4. Hasil penelitian diharapkan dapat membantu pemerintah untuk bertindak cepat dalam mengambil keputusan bagi daerah yang rawan terjangkau penyakit DBD di Kecamatan Setia Janji.

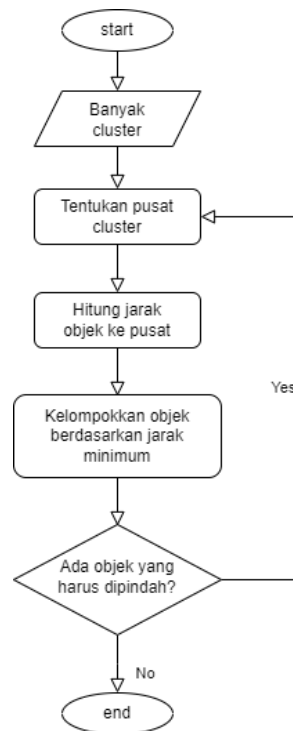
Berdasarkan kedua penelitian tersebut dapat mendukung penelitian yang akan dilakukan berkaitan dengan *clustering* Kabupaten/Kota di Provinsi Jawa Tengah tahun 2022 berdasarkan jumlah kasus kemunculan penyakit dengan algoritma K-Means (Franceschi, 2018; Hashmi, 2020; Jia, 2018; Nguyen, 2019; Udler, 2018). Penelitian ini bertujuan untuk melakukan *clustering* atau pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit-penyakit pada tahun 2022. Dalam penelitian ini, digunakan berbagai data terkait jumlah kasus kemunculan penyakit pada Kabupaten/Kota di Provinsi Jawa Tengah tahun 2022. Dengan menerapkan algoritma K-Means pada data jumlah kasus penyakit tertentu di masing-masing Kabupaten/Kota, kemudian dapat mengelompokkan Kabupaten/Kota tersebut menjadi beberapa kelompok berdasarkan tingkat *cluster* jumlah kasus tersebut. Analisis *clustering* ini diharapkan dapat memberikan wawasan dan pemahaman yang lebih mendalam terkait pola penyebaran penyakit tertentu di wilayah tersebut dan dapat menjadi landasan untuk pengambilan keputusan yang lebih efektif dalam upaya penanganan dan pengendalian penyakit tertentu di Provinsi Jawa Tengah pada masa ke depannya.

Metode

Dalam pemanfaatan algoritma *clustering K-Means* untuk mengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit, diperlukan data terkait hal tersebut (Guo, 2021; Islam, 2018; Larijani, 2019). Sumber data penelitian diperoleh dari data yang dipublikasikan oleh Dinas Kesehatan Provinsi Jawa Tengah pada Buku Saku Kesehatan Provinsi Jawa Tengah Triwulan 1 Tahun 2023. Data yang digunakan dalam penelitian ini adalah data jumlah kasus kemunculan penyakit tertentu pada tahun 2022 di setiap Kabupaten/Kota Provinsi Jawa Tengah yang terdiri dari 29 kabupaten dan 6 kota. Variabel yang digunakan ada 11, yang terdiri dari jumlah kasus kemunculan penyakit campak, gangguan kesehatan mental dan emosional, tuberkulosis, hiv, aids, kusta, filariasis, malaria, leptospirosis, pneumonia, dan ispa.

Data yang telah tersedia akan diolah menggunakan *software* Python. Untuk proses *clustering*, akan menggunakan *package* pandas, matplotlib, seaborn, numpy, dan scikit-learn (sklearn), kemudian untuk visualisasi peta akan menggunakan *package* geopandas dengan data geografinya diambil dari <http://www.diva-gis.org/gdata>. Data akan diolah dengan melakukan *clustering* dalam 3 *cluster*, yaitu *cluster* jumlah kasus kemunculan penyakit tinggi, *cluster* jumlah kasus kemunculan penyakit sedang, dan *cluster* jumlah kasus kemunculan penyakit rendah. Dalam menentukan *cluster* berdasarkan data yang telah tersedia, dibutuhkan sebuah *flowchart* untuk memudahkan dalam menemukan alur

perhitungan. Berikut adalah *flowchart* dalam menentukan *cluster* dengan algoritma *K-Means* menggunakan *software* Python (Sadewo et al., 2017).



Gambar 1. *Flowchart Tahapan Clustering*

Dalam menentukan nilai pusat *cluster* untuk awal iterasi, nilai awal pusat *cluster* dilakukan secara acak, sedangkan jika menentukan pusat *cluster* yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut:

$$\bar{v}_{ij} = \frac{1}{N} \sum_{k=0}^{N_i} x_{kj} \quad (1)$$

1. Menghitung jarak antara titik pusat *cluster* dengan titik tiap objek.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

2. Mengelompokkan objek dengan memperhitungkan jarak minimum objek.
3. Kembali ke langkah ke-2, lakukan iterasi hingga nilai yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.

Hasil dan Pembahasan

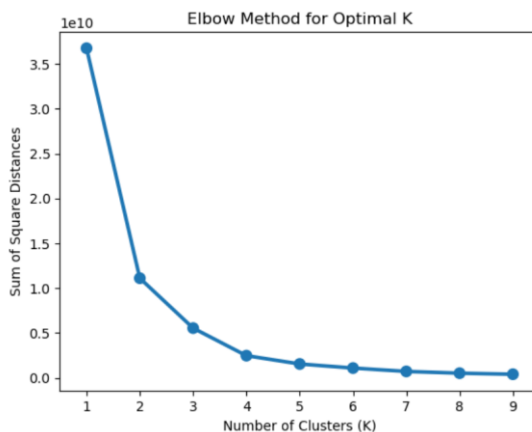
Dalam penelitian *clustering* Kabupaten/Kota di Provinsi Jawa Tengah tahun 2022 berdasarkan jumlah kasus kemunculan penyakit dengan algoritma K-Means, data yang diperoleh berdasarkan Dinas Kesehatan Provinsi Jawa Tengah pada Buku Saku Kesehatan Provinsi Jawa Tengah Triwulan 1 Tahun 2023 melalui situs <https://dinkesjatengprov.go.id/v2018/buku-saku-2/> dapat ditabulasikan ke dalam Tabel 1 berikut.

Tabel 1. Data Kabupaten/Kota Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit

Id	Campak	GME	TB	HIV	AIDS	Kusta	Filariasis	Malaria	Leptospirosis	Pneumonia	Ispa
Cilacap	113	313	98	131	49	31	1	24	26	5847	99785
Banyumas	114	726	151	39	91	17	0	20	73	2098	44074
Purbalingga	24	1055	35	48	50	12	0	12	0	3643	74899
Banjarnegara	10	94	31	24	47	3	0	14	12	3267	53767
Kebumen	21	671	62	79	57	0	0	3	52	5151	55316
Purworejo	65	944	32	54	25	8	1	561	21	61	1412
Wonosobo	83	8171	21	50	23	8	1	18	0	2504	23117
Magelang	135	2803	9	42	22	5	0	17	6	268	3503
Boyolali	76	444	19	129	15	17	2	13	20	3243	47416
Klaten	56	1642	36	100	42	17	0	7	80	3641	70726
Sukoharjo	202	1051	44	64	33	21	0	17	24	1120	29255
Wonogiri	13	64	25	69	44	9	1	7	21	321	36608
Karanganyar	73	3300	23	126	28	7	1	2	23	4052	47862
Sragen	27	126	14	135	30	23	0	72	33	238	38649
Grobogan	29	423	34	220	14	21	1	12	5	1655	13772
Blora	63	433	23	155	63	82	1	24	0	1489	28298
Rembang	28	1855	25	49	34	13	1	5	0	415	7569
Pati	101	844	58	112	28	49	0	38	28	8121	82498
Kudus	18	2504	145	78	81	17	1	5	0	55	2994
Jepara	66	0	51	157	40	71	0	7	4	4275	85648
Demak	27	35	45	106	20	20	0	17	42	5062	23542
Semarang	43	614	11	90	25	12	0	6	3	1476	45425
Temanggung	79	605	20	32	20	1	0	2	0	1149	15174
Kendal	26	461	38	111	41	7	0	6	0	362	19266
Batang	15	0	17	105	41	34	1	3	0	0	16586
Pekalongan	0	5	57	34	7	61	0	0	0	815	6555
Pemalang	14	288	70	139	7	99	1	2	0	2279	79196
Tegal	37	34	133	110	40	120	0	19	0	5969	80185

Brebes	52	861	100	81	68	117	3	1	0	4826	68171
Kota Magelang	1	0	28	9	4	0	0	0	30	389	10532
Kota Surakarta	11	2292	49	23	77	3	0	15	0	234	35361
Kota Salatiga	22	2292	19	31	70	1	0	6	0	1185	8459
Kota Pekalongan	8	859	28	48	23	41	0	2	0	314	5572
Kota Tegal	9	0	80	29	10	8	0	0	0	546	31992
Kota Semarang	74	0	111	311	40	12	0	68	0	11172	138597

Untuk mengetahui jumlah *cluster* yang optimal untuk mengelompokkan data jumlah kasus kemunculan penyakit pada Tabel 1, akan digunakan metode *Elbow*. Metode *Elbow* menghitung nilai selisih penurunan nilai *Sum of Square Error* (SSE). Grafik selisih penurunan nilai *Sum of Square Error* (SSE) dapat dilihat pada gambar berikut.



Gambar 2. Penerapan *Elbow Method* untuk menentukan jumlah *cluster* yang optimal

Dari Gambar 2 di atas, terlihat ada perubahan arah yang signifikan dari garis antara angka 2 dan 3 dengan garis antara 3 dan 4. Garis antara angka 2 dan 3 lebih mengarah ke bawah, sedangkan garis antara angka 3 dan 4 lebih mengarah ke kanan, sehingga jumlah *cluster* yang optimal untuk meng-*cluster* data jumlah kasus kemunculan penyakit dalam penelitian ini adalah 3 *cluster*.

Setelah diketahui jumlah *cluster* yang optimal, yaitu 3, selanjutnya akan ditentukan pusat *cluster*-nya. Berikut adalah pusat dari *cluster* 1, *cluster* 2, dan *cluster* 3.

Tabel 2. Pusat *cluster*

Cluster	Ampanak	ME	B	IV	IDS	usta	ilarias	alari	ptospir	neum	I
1	5 9.67	5 59.67	5 6.89	32.1 1	0.44	8.67	0 .55	1 9.77	15 .33	5 530.33	8 6633.89
2	5 5.17	8 17.92	4.33	0.02	3.33	6.83	0 .42	1 6.08	21 .75	1 936.25	4 1168.58
3	3 8.29	1 496.93	7	9.21	0.36	6.93	0 .43	4 6.71	7 43	1 016.71	1 1289.5

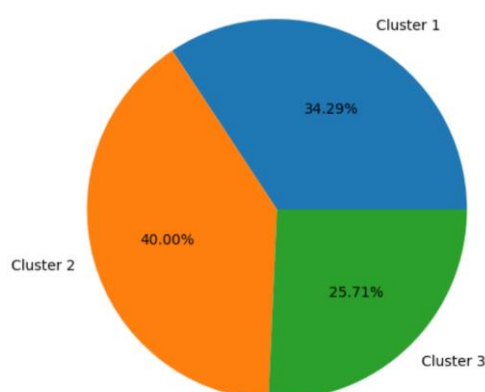
Dari Tabel 2, dapat diketahui bahwa *cluster* 1 adalah *cluster* dengan jumlah kasus kemunculan penyakit tinggi, *cluster* 2 adalah *cluster* dengan jumlah kasus kemunculan penyakit sedang, dan *cluster* 3 adalah *cluster* dengan jumlah kasus kemunculan penyakit

rendah. Setelah pusat *cluster* diketahui, akan dilakukan perhitungan jarak setiap objek dengan pusat *cluster*. Kemudian, hasil perhitungan akan dikelompokkan berdasarkan jarak minimum. Perhitungan ini dilakukan dalam beberapa iterasi sampai tidak ada objek yang perlu dipindah. Berikut adalah hasil *clustering* Kabupaten/Kota Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit dalam *cluster 1*, *cluster 2*, atau *cluster 3*.

Tabel 3. Hasil *Clustering* Kabupaten/Kota Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit

Id	Cluster	Id	Cluster
Cilacap	3	Kudus	2
Banyumas	1	Jepara	3
Purbalingga	3	Demak	2
Banjarnegara	1	Semarang	1
Kebumen	1	Temanggung	2
Purworejo	2	Kendal	2
Wonosobo	2	Batang	2
Magelang	2	Pekalongan	2
Boyolali	1	Pemalang	3
Klaten	3	Tegal	3
Sukoharjo	1	Brebes	3
Wonogiri	1	Kota Magelang	2
Karanganyar	1	Kota Surakarta	1
Sragen	1	Kota Salatiga	2
Grobogan	2	Kota Pekalongan	2
Blora	1	Kota Tegal	1
Rembang	2	Kota Semarang	3
Pati	3		

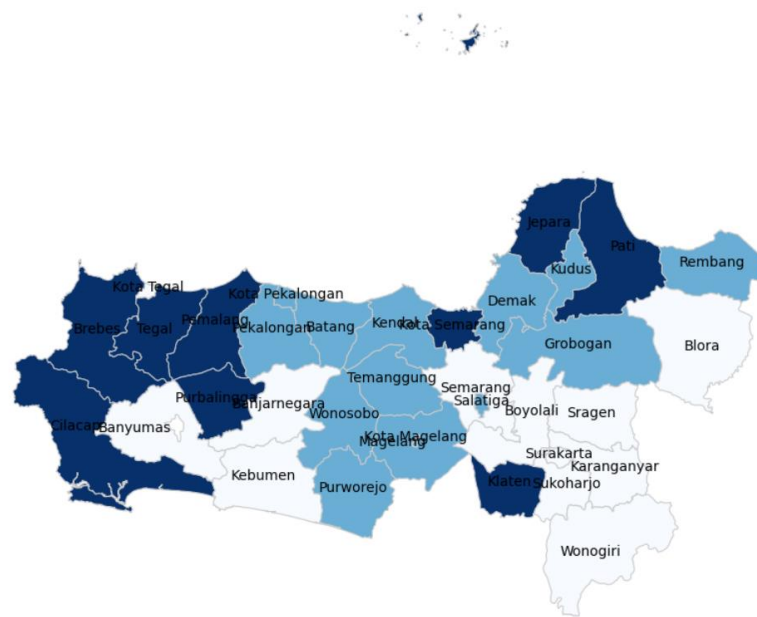
Dari Tabel 3, dapat diketahui ada sebanyak 10 kabupaten dan 2 kota yang masuk *cluster 1*, sebanyak 11 kabupaten dan 3 kota yang masuk *cluster 2*, sebanyak 8 kabupaten dan 1 kota yang masuk *cluster 3*. Persentase hasil *cluster* ini dapat disajikan dalam bentuk *pie chart* sebagai berikut.



Gambar 3. Persentase hasil *Cluster* Kabupaten/Kota Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit

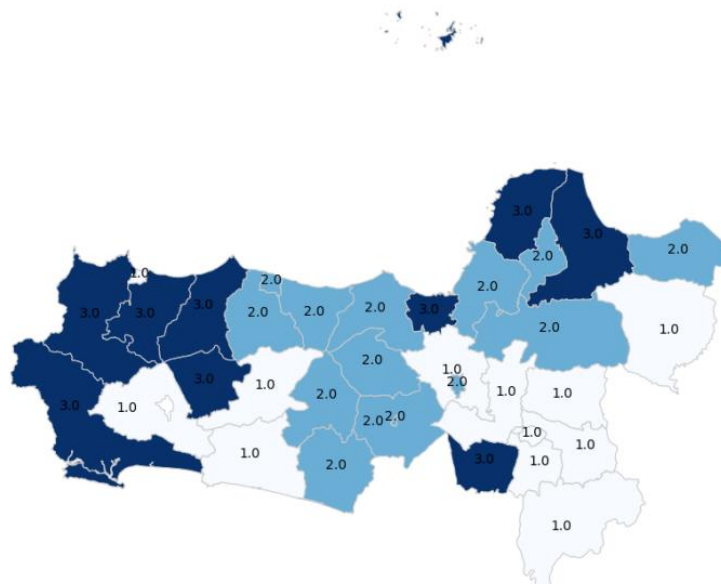
Berdasarkan Gambar 3 di atas dapat dianalisis bahwa persentase untuk *cluster* 1 yang mendefinisikan hasil dari jumlah kasus kemunculan penyakit tinggi memiliki persentase sebesar 34,29%, kemudian untuk *cluster* 2 dengan kategori sedang memiliki persentase sebesar 40,00%, sedangkan untuk *cluster* 3 dengan kategori rendah memiliki persentase sebesar 25,71%. Persentase untuk *cluster* 1 yaitu tinggi dengan perolehan persentase sebesar 34,29% sudah tergolong sebagai perolehan Kabupaten/Kota di Provinsi Jawa Tengah dan wajib menjadi perhatian pemerintah. Namun demikian, Kabupaten/Kota yang masuk dalam *cluster* 2 dan *cluster* 3 juga tetap perlu adanya tindakan dari pemerintah agar tidak mencapai dan termasuk ke dalam kategori *cluster* 1 atau tinggi dalam hal jumlah kasus kemunculan penyakit tertentu.

Kemudian terkait dengan pemetaan hasil *clustering* yang diperoleh dari Tabel 3 di atas, dapat dilakukan pemetaan Kabupaten/Kota dengan pengelompokan *cluster* yang sama melalui pemberian warna pada peta Provinsi Jawa Tengah dengan tiap Kabupaten/Kota masing-masing. Representasi warna yang sama nantinya akan menandakan kesamaan *cluster* tiap Kabupaten/Kota. Di mana untuk *cluster* 1 akan diberi warna putih, *cluster* 2 diberi warna biru muda, dan *cluster* 3 diberi warna biru tua. Berikut hasil peta *cluster* jumlah kasus kemunculan penyakit di Provinsi Jawa Tengah tertampil pada Gambar 4 di bawah.



Gambar 4. Peta *Cluster* Jumlah Kasus Kemunculan Penyakit di Provinsi Jawa Tengah.

Dari hasil visualisasi pemetaan *cluster* jumlah kasus kemunculan penyakit di masing-masing Kabupaten/Kota di Provinsi Jawa Tengah di atas, dapat diidentifikasi dalam pemberian label Kabupaten/Kota mana yang termasuk *cluster* 1 dengan memberi penomoran 1.0, *cluster* 2 dengan 2.0 dan *cluster* 3 dengan 3.0. Dapat dilihat pada Gambar 5 di bawah terkait hasil identifikasi tersebut.



Gambar 5. Identifikasi 3 *Cluster* Kabupaten/Kota Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit

Simpulan

Berdasarkan hasil dan pembahasan terkait *clustering* di Provinsi Jawa Tengah tahun 2022 berdasarkan jumlah kasus kemunculan penyakit tertentu menggunakan algoritma *K-Means* dengan bantuan *software* Python dan *package* pendukung dapat diperoleh hasil 3 *cluster* yaitu tinggi, sedang, dan rendah untuk 29 Kabupaten dan 6 Kota. Persentase untuk *cluster* 1 sebesar 34,29% yang terdiri atas 10 kabupaten dan 2 kota, *cluster* 2 sebesar 40,00% yang terdiri atas 11 kabupaten dan 3 kota, dan *cluster* 3 sebesar 25,71% yang terdiri atas 8 kabupaten dan 1 kota.

Dari hasil tersebut dapat disimpulkan bahwasanya *clustering* Kabupaten/Kota Provinsi Jawa Tengah berdasarkan jumlah kasus kemunculan penyakit tertentu dapat memberikan manfaat bagi pemerintah Provinsi Jawa Tengah karena dapat memberikan pemahaman dalam menyikapi pola persebaran dan tren kemunculan penyakit tertentu melalui visualisasi pemetaan yang dihasilkan. Dengan demikian pemerintah dapat melakukan pendekatan dalam penanganan maupun pencegahan yang berbeda untuk setiap kelompok *cluster*, seperti pengembangan program pencegahan, peningkatan fasilitas kesehatan, atau alokasi sumber daya yang lebih efisien agar Provinsi Jawa Tengah kedepannya mampu menurunkan bahkan menghilangkan angka atau jumlah kemunculan penyakit tertentu dan tercapai pembangunan kesehatan yang baik untuk masyarakatnya.

Daftar Pustaka

- A. S. Wibowo and I. D. Mulyastuti, "Penerapan Algoritma K-Means Clustering Pada Jumlah Fasilitas Kesehatan Menurut Pemerintah Provinsi DKI Jakarta," 2022. [Online]. Available: <https://jakarta.bps.go.id>
- Abualigah, L. M. (2018). Hybrid clustering analysis using improved krill herd algorithm. *Applied Intelligence*, 48(11), 4047–4071. <https://doi.org/10.1007/s10489-018-1190-6>
- Afzal, A. (2020). Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows. *Aerospace Science and Technology*, 107. <https://doi.org/10.1016/j.ast.2020.106318>
- Alam, M. S. (2019). Automatic human brain tumor detection in mri image using template-based k means and improved fuzzy c means clustering algorithm. *Big Data and Cognitive Computing*, 3(2), 1–18. <https://doi.org/10.3390/bdcc3020027>
- Borlea, I. D. (2021). A Unified Form of Fuzzy C-Means and K-Means algorithms and its Partitional Implementation. *Knowledge-Based Systems*, 214. <https://doi.org/10.1016/j.knosys.2020.106731>
- Dinas Kesehatan Provinsi Jawa Tengah, "Buku Saku Kesehatan Provinsi Jawa Tengah Triwulan 1 Tahun 2023," 2023. [Online]. Available: <https://dinkesjatengprov.go.id/v2018/buku-saku-2/>

- Dubey, A. (2018). Comparative study of K-means and fuzzy C-means algorithms on the breast cancer data. *International Journal on Advanced Science, Engineering and Information Technology*, 8(1), 18–29. <https://doi.org/10.18517/ijaseit.8.1.3490>
- Franceschi, F. (2018). Discovering relationships and forecasting PM₁₀ and PM_{2.5} concentrations in Bogotá Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering. *Atmospheric Pollution Research*, 9(5), 912–922. <https://doi.org/10.1016/j.apr.2018.02.006>
- Gianniou, P. (2018). Clustering-based analysis for residential district heating data. *Energy Conversion and Management*, 165, 840–850. <https://doi.org/10.1016/j.enconman.2018.03.015>
- Guo, Z. (2021). Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. *Geoscience Frontiers*, 12(6). <https://doi.org/10.1016/j.gsf.2021.101249>
- Hashmi, M. R. (2020). m-Polar Neutrosophic Topology with Applications to Multi-criteria Decision-Making in Medical Diagnosis and Clustering Analysis. *International Journal of Fuzzy Systems*, 22(1), 273–292. <https://doi.org/10.1007/s40815-019-00763-2>
- Islam, M. (2018). Combining K-MEANS and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering. *Expert Systems with Applications*, 91, 402–417. <https://doi.org/10.1016/j.eswa.2017.09.005>
- Ismkhan, H. (2018). I-k-means+: An iterative clustering algorithm based on an enhanced version of the k-means. *Pattern Recognition*, 79, 402–413. <https://doi.org/10.1016/j.patcog.2018.02.015>
- Jia, C. (2018). O-GlcNAcPRED-II: An integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique. *Bioinformatics*, 34(12), 2029–2036. <https://doi.org/10.1093/bioinformatics/bty039>
- K. D. Sitinjak, A. B. Pangestu, and N. B. Sari, “Clustering Jumlah Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma K-Means,” *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, pp. 46–54, 2022.
- Larijani, M. R. (2019). Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means. *Food Science and Nutrition*, 7(12), 3922–3930. <https://doi.org/10.1002/fsn3.1251>
- M. A. Sembiring, T. A. Agus, M. Fitri, and L. Sibuea, “Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (DBD),” 2021. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>

- M. G. Sadewo, A. P. Windarto, S. R. Andani, and Handrizal, "Pemanfaatan Algoritma Clustering Dalam Mengelompokkan Jumlah Desa / Kelurahan Yang Memiliki Sarana Kesehatan Menurut Provinsi Dengan K-Means," KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer), vol. 1, no. 1, pp. 124–131, 2017, [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/index.php/komik/article/view/483/424>
- Nguyen, H. (2019). A new soft computing model for estimating and controlling blast-produced ground vibration based on Hierarchical K-means clustering and Cubist algorithms. *Applied Soft Computing*, 77, 376–386. <https://doi.org/10.1016/j.asoc.2019.01.042>
- R. Efendi, F. Farady Coastera, and F. Rizky Tanjung, "Pengelompokan dan Pemetaan Derajat Kesehatan Kota Bengkulu dengan Metode K-Means Clustering," 2019. [Online]. Available: <http://ejournal.unib.ac.id/index>
- Rezaee, M. J. (2021). GBK-means clustering algorithm: An improvement to the K-means algorithm based on the bargaining game. *Knowledge-Based Systems*, 213. <https://doi.org/10.1016/j.knosys.2020.106672>
- Sinaga, K. P. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Sun, M. (2019). Pedestrian crash analysis with latent class clustering method. *Accident Analysis and Prevention*, 124, 50–57. <https://doi.org/10.1016/j.aap.2018.12.016>
- Udler, M. S. (2018). Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Medicine*, 15(9). <https://doi.org/10.1371/journal.pmed.1002654>
- Xie, H. (2019). Improving K-means clustering with enhanced Firefly Algorithms. *Applied Soft Computing Journal*, 84. <https://doi.org/10.1016/j.asoc.2019.105763>
- Yang, M. S. (2019). A feature-reduction multi-view k-means clustering algorithm. *IEEE Access*, 7, 114472–114486. <https://doi.org/10.1109/ACCESS.2019.2934179>
- Yu, S. S. (2018). Two improved k-means algorithms. *Applied Soft Computing Journal*, 68, 747–755. <https://doi.org/10.1016/j.asoc.2017.08.032>
- Zhang, G. (2018). Improved K-means algorithm based on density Canopy. *Knowledge-Based Systems*, 145, 289–297. <https://doi.org/10.1016/j.knosys.2018.01.031>
- Zhao, S. (2018). A user-adaptive algorithm for activity recognition based on K-means clustering, local outlier factor, and multivariate gaussian distribution. *Sensors (Switzerland)*, 18(6). <https://doi.org/10.3390/s18061850>

Zhu, E. (2019). Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index. *Neurocomputing*, 363, 149–170. <https://doi.org/10.1016/j.neucom.2019.07.048>