OPEN ACCESS

Volume 9 (1), January-March 2025, 52-58

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v9i1.3007

Analisis *Clustering* Penyakit Menular pada Manusia di Jakarta Timur Menggunakan Algoritma K-Means

Muhammad Arya Ramadhan [1*], Edhy Poerwandono [2], Yuma Akbar [3], Aditya Zakaria Hidayat [4]

[1,2,3,4] Program Studi Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibu Kota Jakarta, Indonesia.*

a r t i c l e i n f o

Article history:
Received 30 July 2024
Received in revised form 25 August 2024
Accepted 25 October 2024
Available online January 2025.

Keywords:
K-Means Algorithm; Clustering; Infectious Diseases; East Jakarta; Data Analysis.

a b s t r a c t

Humans are highly susceptible to various diseases without realizing their causes. The high incidence of infectious diseases in East Jakarta requires an analysis of distribution patterns to determine intervention priorities. This study aims to identify clusters of infectious diseases in East Jakarta, helping authorities plan effective prevention and treatment strategies. Data on infectious disease cases were obtained from the Central Statistics Agency of DKI Jakarta. The K-Means algorithm was used to cluster data based on variables such as period, region, type of disease, and number of cases. The results indicate several main clusters with distinct characteristics that can serve as a foundation for targeted strategies. From 2018 to 2021, diarrhea was predominant, making up 84.14% of cases in 2018 and 81.97% in 2019, pneumonia accounted for 32.92% in 2020, and TB Paru 33.63% in 2021. In conclusion, the K-Means algorithm effectively clusters infectious disease data and provides useful insights into disease distribution in East Jakarta, improving the impact of data-driven health programs.

Kata Kunci:
Algoritma K-Means; Clustering; Penyakit Menular; Jakarta Timur; Analisis Data.

a b s t r a k

Manusia sangat rentan terhadap berbagai penyakit tanpa menyadari penyebabnya. Tingginya kasus penyakit menular di Jakarta Timur memerlukan analisis pola sebaran untuk menentukan prioritas intervensi. Penelitian ini bertujuan mengidentifikasi kelompok penyakit menular di Jakarta Timur, membantu pihak berwenang dalam merencanakan strategi pencegahan dan penanganan yang efektif. Data kasus penyakit menular diperoleh dari Badan Pusat Statistik DKI Jakarta. Metode yang digunakan adalah algoritma K-Means untuk mengelompokkan data berdasarkan variabel seperti periode, wilayah, jenis penyakit, dan jumlah kasus. Hasil penelitian menunjukkan beberapa cluster utama dengan karakteristik berbeda yang dapat menjadi dasar strategi penanganan. Dari tahun 2018 hingga 2021, penyakit diare mendominasi dengan 84,14% kasus pada 2018 dan 81,97% pada 2019, pneumonia 32,92% pada 2020, dan TB Paru 33,63% pada 2021. Kesimpulannya, algoritma K-Means efektif dalam mengelompokkan data penyakit menular dan memberikan gambaran baru tentang distribusi penyakit di Jakarta Timur, meningkatkan efektivitas program kesehatan berbasis data.

ACM Computing Classification System (CCS)
EBSCOhost
Communication and Mass Media Complete (CMMC)

Corresponding Author. Email: aryarmdhn198@gmail.com [1].

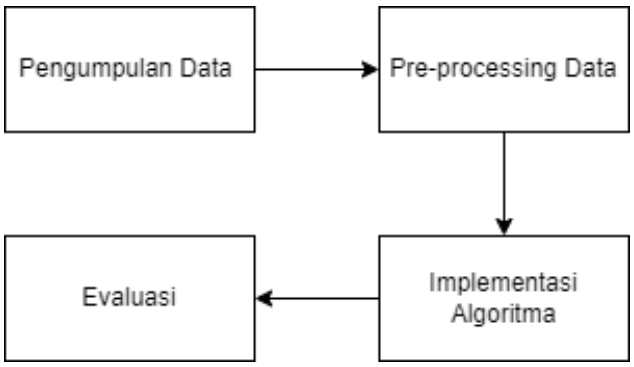
1. Pendahuluan

Penyakit menular telah lama menjadi tantangan utama bagi kesehatan masyarakat di seluruh dunia. Penyebarannya yang cepat, terutama di daerah berpenduduk padat, menjadikan pengendalian penyakit menular sebagai prioritas bagi banyak negara. Di kota besar seperti Jakarta, faktor-faktor seperti urbanisasi, mobilitas tinggi, dan kondisi sanitasi yang buruk memperparah masalah ini. Menurut Nelson *et al.* (2023), peningkatan populasi perkotaan sering dikaitkan dengan tingginya kejadian penyakit menular karena lingkungan padat dan infrastruktur kesehatan yang kurang memadai. Penyakit menular tetap menjadi ancaman besar bagi kesehatan masyarakat di seluruh dunia, terutama di kalangan remaja berusia 10 hingga 19 tahun. Fase transisi kehidupan yang ditandai dengan pertumbuhan cepat membuat remaja lebih rentan terhadap penyakit menular (Nelson *et al.*, 2023).

Di Indonesia, khususnya di Jakarta Timur, penyakit menular seperti demam berdarah dengue (DBD), diabetes, tuberkulosis (TB), malaria, dan lainnya terus menjadi ancaman serius bagi kesehatan masyarakat. Data dari Badan Pusat Statistik DKI Jakarta menunjukkan peningkatan kasus penyakit menular seiring dengan bertambahnya jumlah penduduk dan perkembangan kota yang pesat. Teguh dan Ahmad (2019) mencatat bahwa urbanisasi yang tidak terencana sering menciptakan lingkungan yang mendukung perkembangan vektor penyakit seperti nyamuk *Aedes aegypti*, penyebab utama DBD. Selain itu, kasus diare meningkat akibat kurangnya kesadaran masyarakat akan pentingnya kesehatan. Pada 2017, Provinsi DKI Jakarta mencatat 280.104 kasus diare di fasilitas kesehatan, dengan 250.234 di antaranya ditangani, yang berarti 89,3% dari total kasus (Santos, 2019).

Untuk menghadapi permasalahan ini, diperlukan metode yang efektif untuk memantau dan mengendalikan penyebaran penyakit. Analisis clustering merupakan salah satu pendekatan yang dapat membantu memahami distribusi penyakit menular. Dengan mengelompokkan data berdasarkan karakteristik yang serupa, analisis clustering dapat mengidentifikasi pola penyebaran dan *hotspot* penyakit yang memerlukan perhatian khusus. Algoritma K-Means memungkinkan pengelompokan data berdasarkan kesamaan karakteristik, membantu dalam identifikasi pola penyebaran dan area berisiko tinggi (Sari & Sukestiyarno, 2021). Penelitian ini bertujuan untuk menganalisis *clustering* penyakit menular di Jakarta Timur menggunakan algoritma K-Means. Pengelompokan berdasarkan insiden penyakit dan faktor wilayah dapat membantu mengidentifikasi area berisiko tinggi yang memerlukan intervensi khusus.

2. Metodologi Penelitian



Gambar 1. Tahapan Metodologi

Pengumpulan Data

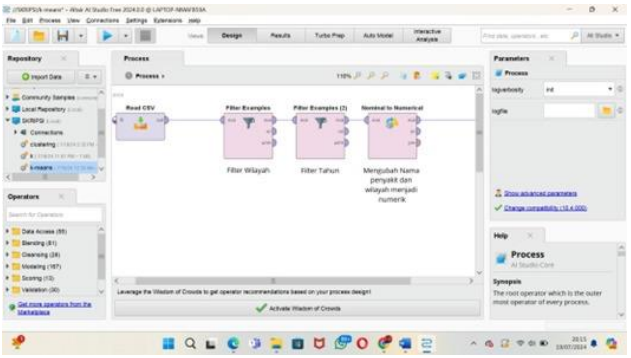
Dataset yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS) dan merupakan dataset publik. Dataset ini berisi data mengenai penderita penyakit menular di wilayah Jakarta Timur. Dalam konteks penelitian ini, dataset tersebut digunakan untuk keperluan clustering, bukan klasifikasi atau segmentasi.

Row No.	wilayah	periode_data	nama_peny...	jumlah
1	Jakarta Timur	2019	TB Paru	12334
2	Jakarta Timur	2019	Pneumonia	3413
3	Jakarta Timur	2019	Kusta	131
4	Jakarta Timur	2019	Campak	330
5	Jakarta Timur	2019	Diare	94495
6	Jakarta Timur	2019	DBD	923
7	Jakarta Timur	2019	AIDS Kasus ...	162
8	Jakarta Timur	2019	AIDS Kasus ...	391
9	Jakarta Timur	2019	IMS	4503

Gambar 2. Data penelitian

Pre-processing Data

Tahap pre-processing melibatkan pembersihan dan transformasi data agar siap digunakan untuk algoritma K-means.

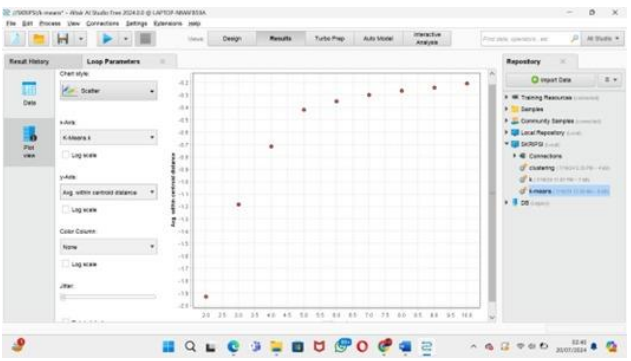


Gambar 3. Pre-processing data

Implementasi Algoritma

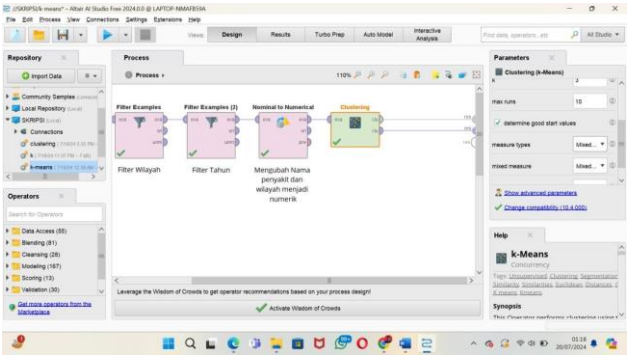
K-means adalah algoritma clustering yang mengelompokkan data ke dalam K cluster berdasarkan jarak terdekat antar data. Untuk menentukan jumlah cluster K yang optimal, metode elbow digunakan. Metode ini melibatkan:

1) Menjalankan K-means untuk berbagai nilai K
2) Menghitung inertia Total jarak kuadrat dari setiap data ke centroid terdekat.
3) Plot inertia terhadap jumlah K Mencari titik "elbow" di mana penurunan inertia mulai melambat, menunjukkan bahwa menambah cluster selanjutnya memberikan keuntungan yang semakin kecil.



Gambar 4. Metode Elbow

Selanjutnya Implementasi K-Means, Melakukan clustering dengan algoritma K-Means pada data yang telah diproses.



Gambar 5. Implementasi K-Means

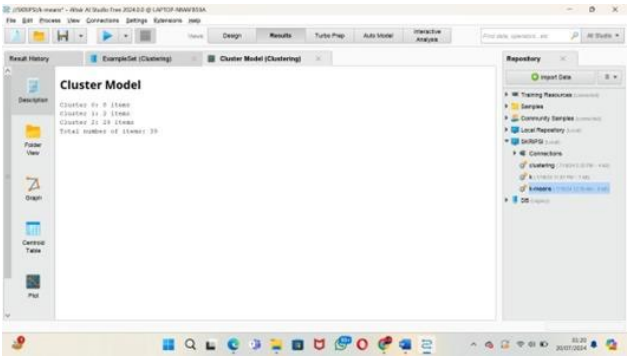
Evaluasi

Dari plot metode elbow di atas, terlihat bahwa titik "elbow" berada di sekitar K = 3. Ini menunjukkan bahwa jumlah cluster yang optimal untuk dataset ini adalah 3.

3. Hasil dan Pembahasan

Hasil

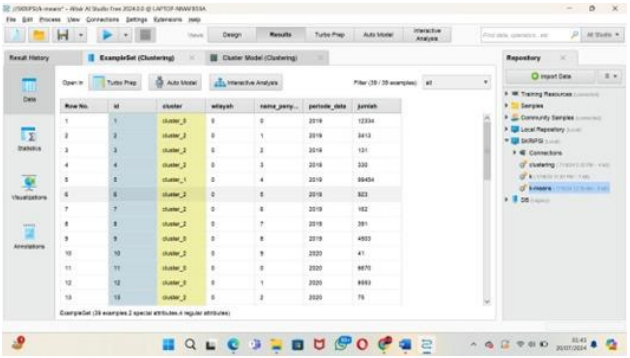
Dalam penelitian ini, proses clustering menggunakan algoritma K-Means menghasilkan tiga cluster dengan karakteristik yang berbeda. Model clustering menunjukkan bahwa Cluster 0 terdiri dari 8 item, Cluster 1 mencakup 2 item, dan Cluster 2 memiliki 29 item. Masing-masing cluster ini memiliki perbedaan dalam jumlah kasus dan karakteristik data yang terkandung di dalamnya.



Gambar 6. Hasil Model Cluster

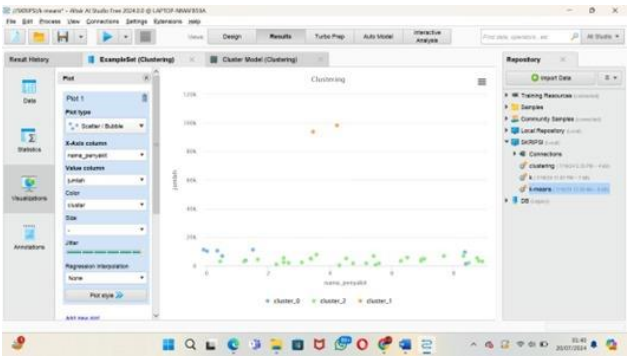
Hasil clustering divisualisasikan untuk memberikan gambaran yang lebih jelas tentang distribusi data. Gambar 6 memperlihatkan model cluster yang dihasilkan oleh algoritma K-Means. Setiap titik dalam visualisasi merepresentasikan satu data poin yang dikelompokkan berdasarkan jarak terdekat dengan

centroid cluster masing-masing. Centroid cluster ditampilkan sebagai titik pusat yang berfungsi sebagai referensi dalam pengelompokan data.



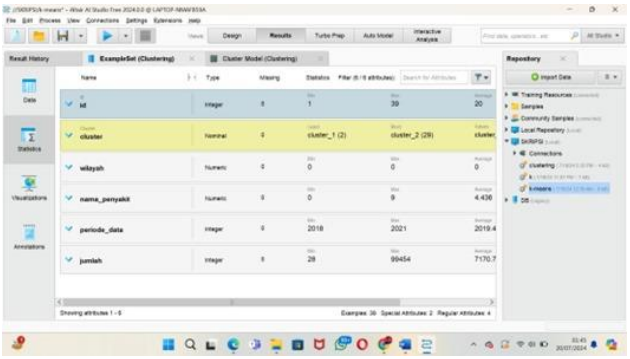
Gambar 7. Hasil Data Clustering

Pengelompokan data dengan K-Means menghasilkan hasil data yang dapat ditinjau lebih lanjut untuk analisis. Gambar 7 menunjukkan hasil data yang telah dikelompokkan ke dalam tiga cluster. Setiap cluster memiliki karakteristik unik berdasarkan variabel-variabel yang digunakan dalam analisis, seperti jenis penyakit, jumlah kasus, periode, dan wilayah.



Gambar 8. Scatter Plot

Gambar 8 menampilkan scatter plot hasil clustering yang memberikan ilustrasi distribusi data dalam ruang multidimensi. Visualisasi ini memungkinkan identifikasi pola dalam data dan melihat bagaimana data dikelompokkan dalam masing-masing cluster. Setiap cluster ditandai dengan warna berbeda untuk mempermudah identifikasi. Scatter plot ini memberikan wawasan mengenai perbedaan signifikan antara cluster satu dengan lainnya dan membantu dalam memahami karakteristik distribusi data. Statistik hasil clustering yang ditampilkan dalam Gambar 9 memberikan informasi detail tentang distribusi data dalam setiap cluster. Cluster 0, yang terdiri dari 8 item, cenderung mewakili kelompok penyakit dengan jumlah kasus yang relatif rendah dibandingkan dengan cluster lainnya. Cluster ini mungkin mencakup penyakit dengan insiden yang jarang terjadi atau penyakit yang memiliki dampak terbatas pada populasi. Cluster 1, yang hanya terdiri dari 2 item, mengindikasikan adanya kelompok penyakit dengan jumlah kasus yang sangat tinggi atau spesifik. Cluster ini bisa mencerminkan jenis penyakit tertentu yang memerlukan perhatian lebih karena penyebarannya signifikan di Jakarta Timur. Berdasarkan penelitian ini, cluster ini bisa berhubungan dengan penyakit seperti diare yang memiliki jumlah kasus tinggi dalam beberapa tahun. Cluster 2, dengan 29 item, mencakup kelompok penyakit dengan jumlah kasus yang bervariasi, berada di antara cluster 0 dan cluster 1 dalam hal jumlah kasus. Cluster ini mungkin menunjukkan penyakit menular dengan frekuensi kejadian yang cukup umum namun tidak seintensif yang tercatat dalam Cluster 1.



Gambar 9. Hasil statistik.

Proses evaluasi hasil clustering menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data secara efisien, dengan jumlah cluster optimal pada *K* = 3 yang ditentukan melalui metode *elbow*. Setiap cluster memberikan informasi penting yang dapat digunakan oleh pihak berwenang untuk merumuskan strategi penanganan dan pencegahan yang tepat sasaran. Penyakit dalam Cluster 1 memerlukan perhatian khusus karena jumlah kasus yang signifikan, sehingga perlu alokasi sumber daya yang lebih besar untuk pencegahan dan penanganan. Cluster 0 dan 2 memerlukan strategi penanganan yang berbeda, dengan fokus pada upaya pengendalian dan pencegahan sesuai dengan karakteristik masing-masing cluster. Analisis ini menunjukkan bahwa metode clustering K-Means tidak hanya efektif dalam mengelompokkan data, tetapi juga menyediakan landasan bagi perencanaan intervensi kesehatan yang

lebih strategis dan terarah.

Pembahasan

Analisis *clustering* menggunakan algoritma K-Means menunjukkan bahwa pengelompokan data penyakit menular di Jakarta Timur menghasilkan tiga cluster yang berbeda, masing-masing dengan karakteristik khusus. Penggunaan metode ini membantu mengidentifikasi distribusi penyakit menular berdasarkan jumlah kasus dan variasi penyebaran di wilayah tersebut. Penelitian ini sejalan dengan temuan dari Lessler *et al.* (2016), yang menunjukkan pentingnya mengukur ketergantungan spasial dalam epidemiologi penyakit menular untuk memahami pola distribusi di berbagai area. Cluster 0 mencakup penyakit dengan jumlah kasus yang lebih rendah, mencerminkan kelompok penyakit yang memiliki insiden kecil. Hal ini menunjukkan bahwa meskipun ada kehadiran penyakit di wilayah tersebut, penyebarannya relatif terkendali dibandingkan cluster lainnya. Bhunia *et al.* (2021) menyebutkan pentingnya analisis spasio-temporal dalam memahami dinamika penyebaran penyakit, yang mendukung relevansi penelitian ini dalam mengidentifikasi variasi insiden penyakit menular di berbagai wilayah.

Cluster 1, di sisi lain, mencakup penyakit dengan jumlah kasus yang sangat tinggi. Misalnya, diare yang mencapai angka 99.454 kasus termasuk dalam cluster ini. Temuan ini menyoroti area dengan penyebaran signifikan yang memerlukan perhatian dan intervensi prioritas. Seperti yang diuraikan oleh Qomariasih (2021), penggunaan algoritma K-Means dapat membantu pemerintah dalam menentukan area yang memerlukan penanganan lebih intensif, khususnya dalam konteks pandemi COVID-19 di Jakarta. Implementasi algoritma ini dalam pengelompokan wilayah berdampak langsung pada penyusunan strategi penanganan dan alokasi sumber daya yang tepat. Cluster 2 terdiri dari penyakit dengan jumlah kasus yang berada di antara cluster 0 dan cluster 1. Kelompok ini menunjukkan penyakit dengan insiden cukup umum tetapi tidak seintensif cluster dengan jumlah kasus tinggi. Penelitian oleh Tukiyat dan Djohan (2022) menunjukkan bahwa metode clustering, termasuk K-Means, efektif dalam memetakan penyebaran pandemi COVID-19 di kota besar seperti Jakarta, membantu dalam memahami variasi dan penyusunan strategi penanggulangan.

Rustam *et al.* (2018) juga mencatat bahwa optimasi K-Means yang dikombinasikan dengan algoritma *particle swarm optimization* efektif dalam mengidentifikasi daerah endemik, menunjukkan relevansi pendekatan ini untuk berbagai jenis penyakit menular. Penggunaan K-Means diperkuat oleh aplikasi praktis dalam studi lainnya, seperti yang dilakukan oleh Sari *et al.* (2018) yang menggunakan RapidMiner untuk implementasi K-Means dalam data imunisasi campak. Hasil dari implementasi ini menunjukkan bagaimana data dapat dikelompokkan secara efisien untuk menentukan prioritas kampanye kesehatan. Halim *et al.* (2022) menambahkan bahwa algoritma ini mampu mengelompokkan wilayah dengan penyebaran COVID-19 di Indonesia, memberikan gambaran distribusi yang berguna bagi pembuat kebijakan.

Penerapan algoritma K-Means untuk memetakan penyakit menular juga didukung oleh Kurniawan *et al.* (2023), yang menunjukkan bahwa analisis ini berguna untuk mengelompokkan area penyebaran COVID-19 di DKI Jakarta. Hasil clustering ini membantu dalam merancang strategi intervensi berdasarkan tingkat kasus yang ditemukan di setiap cluster. Siringi *et al.* (2020) juga mendemonstrasikan efektivitas K-Means dalam mengidentifikasi hotspot demam berdarah, di mana pengelompokan data memungkinkan penentuan wilayah prioritas untuk pencegahan. Implementasi RapidMiner mempermudah visualisasi hasil clustering, memungkinkan identifikasi pola distribusi yang lebih cepat dan efisien. Silvi (2018) yang mengelompokkan data indikator HIV/AIDS di Indonesia menunjukkan bahwa analisis clustering dapat memberikan hasil signifikan dalam mengarahkan kebijakan kesehatan publik dan strategi pencegahan. Algoritma K-Means terbukti efektif dalam mengelompokkan data penyakit menular di Jakarta Timur. Studi ini mendukung literatur sebelumnya yang menekankan pentingnya analisis clustering dalam pengambilan keputusan terkait kesehatan masyarakat, baik untuk perencanaan intervensi maupun alokasi sumber daya yang lebih tepat

Hasil clustering menggunakan algoritma K-Means menunjukkan bagaimana data dikelompokkan ke dalam tiga cluster yang berbeda, masing-masing

dengan karakteristik jumlah kasus yang serupa. Visualisasi cluster mempermudah identifikasi dan pemahaman mengenai distribusi data serta perbedaan antar cluster. Cluster 0 terdiri dari penyakit dengan jumlah kasus yang relatif rendah, menunjukkan kelompok penyakit yang memiliki insiden lebih kecil dibandingkan dengan cluster lainnya. Cluster 1 mencakup penyakit dengan jumlah kasus yang sangat tinggi, seperti diare yang mencapai jumlah kasus 99.454. Cluster ini menandakan penyakit dengan tingkat penyebaran yang tinggi dan membutuhkan perhatian khusus. Cluster 2 terdiri dari penyakit dengan jumlah kasus yang berada di antara cluster 0 dan cluster 1, mencerminkan kelompok penyakit dengan insiden yang cukup umum namun tidak seintensif yang ditemukan di Cluster 1. Analisis hasil clustering ini memberikan gambaran yang lebih jelas mengenai prioritas penanganan. Cluster 1, yang menunjukkan jumlah kasus yang sangat tinggi, memerlukan alokasi sumber daya yang lebih besar untuk pencegahan dan penanganan yang efektif. Sementara itu, Cluster 0 dan Cluster 2 memerlukan strategi penanganan yang disesuaikan dengan karakteristik jumlah kasus yang lebih rendah. Pendekatan ini memungkinkan perencanaan intervensi yang lebih terarah dan efisien, dengan prioritas yang diberikan pada area dengan tingkat penyebaran tertinggi. Implementasi hasil clustering menggunakan algoritma K-Means dan RapidMiner membantu mengidentifikasi pola dalam data penyakit menular di DKI Jakarta. Dengan membagi data ke dalam tiga cluster, analisis ini memfasilitasi pemahaman distribusi jumlah kasus penyakit. Informasi tersebut dapat digunakan untuk merancang strategi penanganan yang lebih efektif berdasarkan karakteristik setiap cluster, memastikan bahwa upaya pencegahan dan pengendalian lebih tepat sasaran dan efisien.

4. Kesimpulan

Penelitian berhasil mengidentifikasi pola penyebaran penyakit menular di Jakarta Timur dengan menggunakan algoritma K-Means. Algoritma ini terbukti efektif dalam mengelompokkan data berdasarkan variabel seperti periode, wilayah, jenis penyakit, dan jumlah orang yang terkena penyakit, menghasilkan tiga cluster utama dengan karakteristik yang berbeda. Hasil clustering memberikan informasi yang berguna bagi pihak berwenang dalam merencanakan strategi penanganan yang lebih spesifik dan efisien, serta berkontribusi pada peningkatan efektivitas program kesehatan di wilayah tersebut melalui pendekatan berbasis data.

5. Daftar Pustaka

Bhunia, G. S., Roy, S. S., & Shit, P. K. (2021). Spatio-temporal analysis of COVID-19 in India – a geostatistical approach. *Spatial Information Research, 29*(5), 661–672. https://doi.org/10.1007/s41324-020-00376-0

Halim, C., Purnomo, H., & Wahyono, T. (2022). Analisis pengelompokan wilayah penyebaran COVID-19 di Indonesia dengan metode clustering menggunakan algoritma K-Means dan K-Medoids. *INOVTEK Polbeng - Seri Informatika, 7*(2). https://doi.org/10.35314/isi.v7i2.2566

Kurniawan, T., Wisjhnuadji, T., & Hanif, H. (2023). Data mining application for clustering COVID-19 spread areas in DKI Jakarta using the K-Means algorithm. *Jurnal Limits, 20*(1). https://doi.org/10.59134/jlmt.v20i1.325

Lessler, J., Salje, H., Grabowski, M. K., & Cummings, D. A. T. (2016). Measuring spatial dependence for infectious disease epidemiology. *PLOS ONE, 11*(5), e0155249. https://doi.org/10.1371/journal.pone.0155249

Mesquita, C. R., Enk, M. J., & Guimarães, R. J. d. P. S. e. (2021). Spatial analysis studies of endemic diseases for health surveillance: Application of scan statistics for surveillance of tuberculosis among residents of a metropolitan municipality aged 60 years and above. *Ciência & Saúde Coletiva, 26*(suppl 3), 5149–5156. https://doi.org/10.1590/1413-812320212611.3.09132020

Qomariasih, N. (2021). Implementasi K-Means clustering analysis untuk mengelompokkan

kelurahan-kelurahan di DKI Jakarta berdasarkan jumlah positif COVID-19. *Jurnal Syntax Transformation, 2*(7). https://doi.org/10.46799/jst.v2i7.336

Rustam, S., Santoso, H., & Supriyanto, C. (2018). Optimasi K-Means clustering untuk identifikasi daerah endemik penyakit menular dengan algoritma particle swarm optimization di Kota Semarang. *ILKOM Jurnal Ilmiah, 10*(3), 251–259. https://doi.org/10.33096/ILKOM.V10I3.342.251-259

Santos, T. B. (2019). Aplikasi data mining untuk clustering daerah penyebaran penyakit diare di DKI Jakarta menggunakan algoritma K-Means. *Jurnal Ilmiah FIFO, 11*(2), 131. https://doi.org/10.22441/fifo.2019.v11i2.003

Sari, D. N. P., & Sukestiyarno, Y. L. (2021). Analisis cluster dengan metode K-Means pada persebaran kasus COVID-19 berdasarkan provinsi di Indonesia. *Prisma: Prosiding Seminar Nasional Matematika, 4*, 602–610.

Sari, R., Wanto, A., & Windarto, A. (2018). Implementasi RapidMiner dengan metode K-Means (studi kasus: imunisasi campak pada balita berdasarkan provinsi). *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer), 2*(1). https://doi.org/10.30865/KOMIK.V2I1.930

Silvi, R. (2018). Analisis cluster dengan data outlier menggunakan centroid linkage dan K-Means clustering untuk pengelompokkan indikator HIV/AIDS di Indonesia. *Journal of Multimedia, 4*, 22–31. https://doi.org/10.15642/MANTIK.2018.4.1.22-31

Siringi, N., Mala, S., & Rawat, A. (2020). Study of K-Means clustering algorithm for identification of dengue fever hotspots. In *Proceedings of the Conference on Emerging Technologies*, 51–61. https://doi.org/10.1007/978-981-15-1420-3_6

Tukiyat, T., & Djohan, Y. (2022). Analisis penyebaran pandemi COVID-19 di Kota Jakarta menggunakan metode clustering K-Means dan Density-Based Spatial Clustering of Application with Noise (DBSCAN). *Jurnal Informatika*. https://doi.org/10.31294/inf.v9i1.11226

Yan, J., Cougoule, C., Lacroix-Lamandé, S., & Wiedemann, A. (2023). Mapping the scientific output of organoids for modeling animal and human infectious diseases: A bibliometric assessment. *Research Square*. https://doi.org/10.21203/rs.3.rs-3691844/v1.