



Perbandingan pengukuran jarak Euclidean dan Gower pada kluster k-medoids

Comparison analysis of Euclidean and Gower distance measures on k-medoids cluster

Agil Aditya^{*)}, Betha Nurina Sari, Tesa Nur Padilah

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. H.S. Ronggowaluyo, Teluk Jambe Timur, Karawang, Jawa Barat, Indonesia 41361

Cara sitasi: A. Aditya, B. N. Sari, and T.N Padilah, "Perbandingan pengukuran jarak Euclidean dan Gower pada kluster k-medoids," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 1, pp. 1-7, 2021. doi: [10.14710/jtsiskom.2021.13747](https://doi.org/10.14710/jtsiskom.2021.13747), [Online].

Abstract – *K-medoids clustering uses distance measurement to find and classify data that have similarities and inequalities. The distance measurement method selection can affect the clustering performance for a dataset. Several studies use the Euclidean and Gower distance as measurement methods in numerical data clustering. This study aims to compare the performance of the k-medoids clustering on a numerical dataset using the Euclidean and Gower distance. This study used seven numerical datasets and Silhouette, Dunn, and Connectivity indexes in the clustering evaluation. The Euclidean distance is superior in two values of Silhouette and Connectivity indexes so that Euclidean has a good data grouping structure, while the Gower is superior in Dunn index showing that the Gower has better cluster separation compared to Euclidean. This study shows that the Euclidean distance is superior to the Gower in applying the k-medoids algorithm with a numeric dataset.*

Keywords – clustering; data mining; Euclidean; Gower; k-medoids

Abstrak – *Klastering k-medoids menggunakan metode jarak untuk mencari dan mengelompokkan data yang memiliki kesamaan dan ketidaksamaan. Penentuan metode pengukuran jarak adalah hal yang penting karena mempengaruhi performa hasil kluster k-medoids. Beberapa kajian menyatakan bahwa metode Euclidean dan Gower bisa digunakan sebagai metode pengukuran pada klastering dengan data numerik. Penelitian ini bertujuan untuk melakukan perbandingan performa hasil klastering k-medoids pada dataset numerik menggunakan metode Euclidean dan Gower. Penelitian ini menggunakan tujuh dataset numerik dan evaluasi hasil klastering menggunakan nilai Silhouette, Dunn, dan Connectivity. Metode jarak Euclidean unggul pada dua nilai evaluasi Silhouette dan Connectivity yang menunjukkan bahwa Euclidean memiliki struktur*

pengelompokan data yang baik, sedangkan Gower unggul pada satu nilai evaluasi Dunn yang menunjukkan Gower memiliki pemisah antar kluster yang baik dibanding Euclidean. Penelitian ini menunjukkan bahwa metode Euclidean lebih unggul daripada metode Gower pada penerapan algoritma k-medoids dengan dataset bertipe numerik.

Kata kunci – klastering; data mining; Euclidean; Gower; k-medoids

I. PENDAHULUAN

Data mining digunakan untuk mengolah data menjadi suatu informasi dan pengetahuan [1]. Dalam prosesnya, data mining menggunakan teknik statistik, matematika, dan kecerdasan buatan untuk identifikasi dan ekstraksi informasi dan pengetahuan. Salah satu teknik dalam data mining yang bisa mengelompokkan data besar untuk memudahkan pencarian informasi yang relevan sesuai dengan kebutuhan adalah klastering.

Klastering merupakan suatu metode data mining yang bersifat tanpa arahan untuk mencari dan mengelompokkan data yang memiliki kesamaan dan ketidaksamaan karakteristik antara satu data dengan data yang lain [2]. Beberapa metode klastering yang sering digunakan adalah k-means, improved k-means, fuzzy c-means, DBSCAN, k-medoids (PAM), CLARANS dan fuzzy subtractive [3].

Kamila dkk. [1] melakukan perbandingan algoritme k-medoids dan k-means dengan menggunakan waktu kecepatan pengolahan dan *davies-bouldin index* sebagai parameternya. Penelitian tersebut menghasilkan algoritme k-medoids lebih unggul dibandingkan k-means. Marlina dkk. [4] menerapkan k-medoids dan k-means berdasarkan nilai *k* atau jumlah kluster untuk melakukan pengelompokan sebaran cacat pada anak. Kajian tersebut menunjukkan bahwa penerapan k-medoids lebih baik dibandingkan dengan k-means. Senduk dkk. [5] menerapkan algoritme k-medoids dan k-means dalam klastering wilayah gempa bumi di Indonesia dan menunjukkan bahwa algoritme k-medoids memberikan nilai Silhouette yang lebih baik daripada k-means. Di sisi lain, k-medoids menghasilkan komputasi

^{*)} Penulis korespondensi (Agil Aditya)
Email: agil.16017@student.unsika.ac.id

yang baik [6] dan konsisten dalam berbagai pengaturan data variabel campuran [7].

K-medoids dapat melakukan pengelompokan menggunakan objek yang paling terpusat (medoids) menjadi sejumlah k kluster dari nilai rata-rata objek dalam sebuah kluster. K-medoids juga mengukur jarak di setiap objek non-medoids, dimana objek dengan jarak terdekat dengan masing-masing medoid akan bergabung dengan medoid tersebut [4]. K-medoids menggunakan pengukuran jarak untuk mengelompokkan data berdasarkan tingkat kesamaan atau kemiripan antara data sehingga data-data yang memiliki tingkat kemiripan paling tinggi akan berkumpul menjadi satu kluster.

Pengukuran jarak mempengaruhi hasil akhir klustering [8], salah satunya mempengaruhi hasil klustering k-medoids [9]. Mohibullah dkk. [10] menyatakan bahwa k-medoids dapat menggunakan berbagai metode pengukuran jarak. Performa hasil kluster k-medoids berhubungan erat dengan teknik pengukuran jarak yang digunakan. Beberapa metode pengukuran jarak yang dapat digunakan adalah Euclidean, Manhattan, Chebyshev, Minkowski, dan Gower [7], [11]–[13]. Hal ini menunjukkan bahwa penentuan metode pengukuran jarak adalah hal yang penting dilakukan dalam klustering menggunakan k-medoids.

Penelitian yang membandingkan kinerja metode pengukuran jarak yang berbeda telah banyak dilakukan [7]–[10], [14]–[19]. Budiaji dan Leisch [7] menggunakan 6 dataset yang berbeda untuk melakukan perbandingan pengukuran jarak. Pandit dan Gupta [8] melakukan perbandingan di antara pengukuran jarak berdasarkan domain aplikasi, efisiensi, manfaat dan kelemahan. Dahal [9] menggunakan validasi kluster untuk mengukur perbedaan antara pengukuran jarak yang digunakan. Mohibullah dkk. [10] mengungkapkan metode Euclidean lebih efektif dibandingkan metode pengukuran jarak lainnya dan cocok digunakan untuk dataset berukuran kecil.

Lebih lanjut, Rani dan Sahu [14] membandingkan teknik klustering hirarki dan k-means untuk mengukur kemiripan dalam artikel. Ali dan Massmoudi [15] memaparkan pengukuran Gower memberikan hasil terbaik dibandingkan dengan metrik yang digunakan lainnya. Nishom [16] memaparkan bahwa pengukuran jarak Euclidean menunjukkan hasil yang baik dari pengukuran jarak Manhattan dan Minkowski. Sinwar dan Kaushik [17] melakukan perbandingan metode jarak Euclidean dan Manhattan dengan dua dataset nyata dan sintetik dan menghasilkan bahwa Euclidean menunjukkan performa lebih baik dalam jumlah iterasi. Sunge dkk. [18] menunjukkan performa k-medoids menggunakan metode pengukuran jarak Euclidean dan Manhattan lebih baik daripada Chebyshev. Perbandingan yang dilakukan Fajriah dkk. [19] juga menunjukkan metode Euclidean lebih unggul dari Manhattan pada dataset numerik.

Beberapa kajian tersebut menjadikan metode Euclidean sebagai metode yang standar untuk digunakan [20]. Hal inilah yang memunculkan pertanyaan apakah

metode Euclidean merupakan metode yang paling dapat diandalkan, padahal masih ada opsi metode pengukuran jarak yang lain, di antaranya Manhattan [17], [20] dan Gower [21] yang juga merupakan pengukuran jarak yang populer.

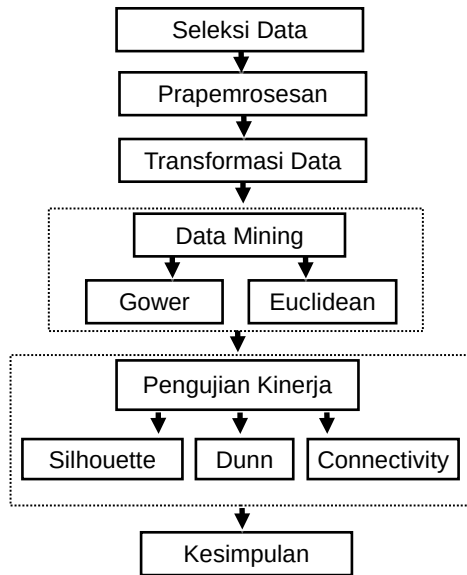
Performa hasil kluster dengan menggunakan metode Euclidean dapat dikatakan baik di metode klustering partisi tradisonal [22]. Hoven [23] menjelaskan bahwa k-medoids dapat digunakan bersamaan dengan metode ketidaksamaan Gower. Jung dkk. [24] menerapkan algoritme k-medoids dan Gower dalam risetnya karena mampu menghilangkan pengaruh *outlier* dan *noise*. Sadowski [25] menggunakan dataset numerik dan metode jarak Euclidean dan Manhattan dalam risetnya dan menyatakan bahwa metode Gower dapat digunakan sebagai alternatif. Nowak-Brzezinska dan Rybotycki [26] menjelaskan metode jarak Gower dapat digunakan untuk menangani atribut numerik dan atribut simbolik berbeda. Gower adalah metode pengukuran jarak yang bisa menangani data campuran, yaitu numerik dan nominal [23]. Euclidean umumnya digunakan untuk pengukuran jarak pada data numerik [15]. Mohammed dan Abdulazeez [27] juga menyatakan bahwa metode Euclidean adalah metrik umum untuk data numerik. Metode Euclidean bekerja secara efisien dengan menghitung kemiripan dalam pengelompokannya dan dapat memisahkan data dengan baik.

Kajian-kajian tersebut menunjukkan bahwa metode Euclidean dan Gower dapat digunakan sebagai metode pengukuran pada klustering dengan data numerik. Namun, belum ada kajian untuk menerapkan dan menjelaskan performa hasil dari metode Gower yang digunakan sebagai metode pengukuran jarak pada klustering k-medoids dengan dataset numerik. Penelitian ini bertujuan untuk melakukan perbandingan performa hasil klustering k-medoids pada dataset numerik menggunakan metode Euclidean dan Gower.

II. METODE PENELITIAN

Metode penelitian yang digunakan adalah *knowledge discovery in database* (KDD) untuk mencari dan mengidentifikasi pola dari data [28]. KDD ini juga dapat digunakan untuk mencari deviasi, analisis kecenderungan, algoritme genetik, jaringan syaraf tiruan dan pendekatan campuran dua atau lebih [29]. Alur penelitian ini ditunjukkan pada [Gambar 1](#).

Urutan pertama pada penelitian ini adalah mengumpulkan dataset. Dataset yang dikumpulkan terdiri dari 7 dataset yang berbeda jumlah objeknya. Pemilihan 7 dataset ini dilakukan untuk membandingkan dua pengukuran jarak yang digunakan dan mengevaluasi performa yang dihasilkan berdasarkan nilai Silhouette, Dunn, dan Connectivity. Dataset yang diperoleh dari repositori UCI, Pusat Penilaian Pendidikan dan Kebudayaan Kementerian Pendidikan dan Kebudayaan Republik Indonesia, Dinas Kesehatan Kabupaten Karawang dan PUSDATIN Kementerian Kesehatan Republik Indonesia ditunjukkan pada [Tabel 1](#).



Gambar 1. Alur penelitian

Pembersihan data yang telah dikumpulkan dilakukan pada tahap pra pemrosesan. Proses ini mencakup di antaranya untuk membuang duplikasi data, memeriksa inkonsisten pada data, menangani data hilang, dan memperbaiki kesalahan pada data seperti kesalahan penulisan. Jika tidak ditemukan data yang inkonsisten pada dataset tersebut, langkah transformasi dilakukan.

Data yang sudah dibersihkan pada langkah sebelumnya kemudian dilakukan transformasi, yaitu untuk mengubah data yang berjenis kategori menjadi numerik untuk mengubah data berbentuk kategori ke numerik. Transformasi juga dilakukan dengan memberikan nilai ordinal kepada setiap kategori pada masing-masing variabel. Standarisasi data dilakukan dengan menggunakan normalisasi *min-max* dengan skala 0-1. Pada proses normalisasi *min-max*, data diubah nilainya ke skala 0 sampai 1. Perangkat lunak yang digunakan dalam proses normalisasi adalah RStudio. Tujuan dari proses ini adalah untuk melakukan penskalaan nilai atribut dari data sehingga bisa jatuh pada jangkauan tertentu.

Metode k-medoids melakukan pengelompokan berdasarkan sekumpulan n objek menjadi sejumlah k kluster. Algoritme ini menggunakan objek pada kumpulan objek yang mewakili sebuah kluster. Objek yang mewakili sebuah cluster disebut dengan medoids. Proses pengelompokan bergantung pada metode jarak yang digunakan. Algoritme k-medoids dapat diringkas mengikuti [Algoritme 1 \[18\]](#).

Metode jarak yang digunakan dalam penelitian ini adalah Euclidean dan Gower. Metode jarak Euclidean digunakan untuk mengukur jarak antara 2 titik yang berbeda dan dinyatakan dalam Persamaan 1 [30]. Metode jarak Gower dihitung sebagai rata-rata perbedaan parsial antar individu. Setiap ketidaksamaan parsial (dan dengan demikian jarak Gower) berkisar dalam [0 1], perhitungan jarak Gower dinyatakan dalam Persamaan 2 [31].

Tabel 1. Dataset yang digunakan

No	Dataset	Objek	Type Data
1	Wholesale customers data	440	Numerik
2	Hepatitis	143	Numerik
3	Contraceptive method choice	1000	Numerik
4	Ujian nasional SMP Indonesia	34	Numerik
5	Penyakit diare Kabupaten Karawang	30	Numerik
6	Penyakit campak perbulan provinsi	34	Numerik
7	Tenaga kesehatan dokter per provinsi	34	Numerik

Algoritme 1. Algoritme k-medoids klustering

Require : k , number of cluster; D , a data set of N points

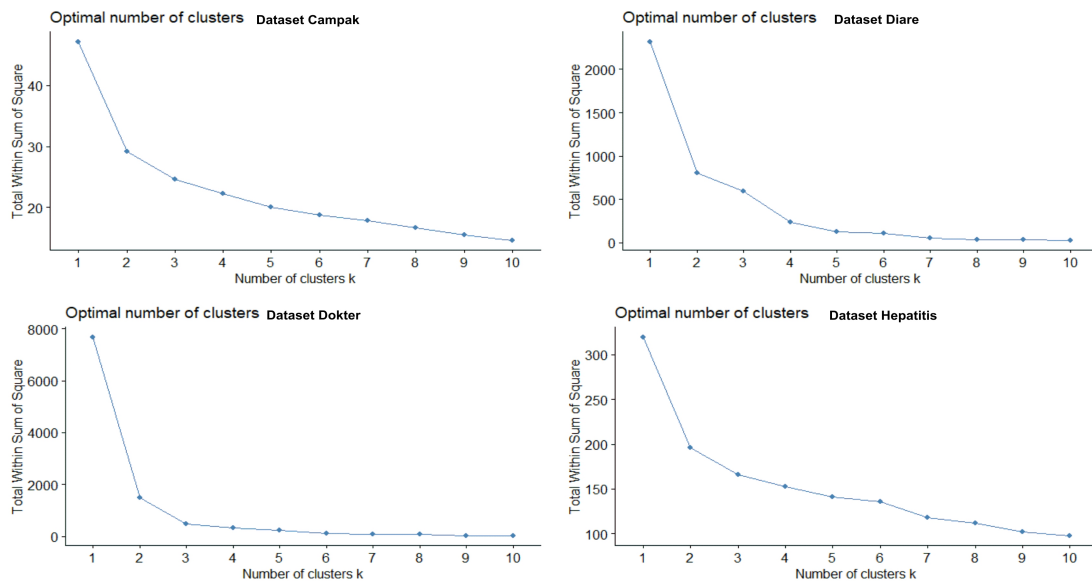
Ensure: A set of k clusters

- 1: Arbitrarily choose k points in D as initial representative points.
- 2: **repeat**
- 3: **for** each non-representative point p in D **do**
- 4: find the nearest representative point and assign p to the corresponding cluster.
- 5: **end for**
- 6: randomly select a non-representative point $prand$
- 7: compute the overall cost C of swapping a representative point pi with $prand$
- 8: **if** $C < 0$ **then**
- 9: swap pi with $prand$ to form a new set of k representative point
- 10: **end if**
- 11: **until** stop-iteration criteria satisfied
- 12: **return** clustering result

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$s(x_i, x_j) = \frac{\sum_{q=1}^l S_Q(x_i, x_j)}{\sum_{q=1}^l W_q} \quad (2)$$

Penelitian ini membandingkan dua metode jarak, yaitu Euclidean dan Gower menggunakan algoritme k-medoids dengan 7 dataset yang berbeda yang dapat dilihat pada [Tabel 1](#). Penentuan nilai k dilakukan dengan menggunakan metode *elbow* dengan bantuan RStudio dengan menggunakan pustaka *factextra*. Jumlah nilai kluster optimal diperoleh pada nilai kluster 2 dan nilai kluster 3 pada tiap dataset berbeda. Hasil proses dari metode *elbow* ditunjukkan pada [Gambar 2](#). Metode *elbow* ini digunakan untuk menghasilkan informasi dalam menentukan jumlah kluster optimal dengan cara melihat



Gambar 2. Metode *elbow* pada dataset campak, dokter, diare, dan hepatitis

persentase hasil perbandingan antara jumlah kluster yang akan membentuk siku pada suatu titik [32], [33].

Setelah itu, proses mining dilakukan dengan menggunakan dua metode jarak yang berbeda. Hasil dari masing-masing dari metode jarak tersebut dievaluasi dengan nilai Silhouette, Dunn, dan Connectivity, seperti yang digunakan dalam [34]-[36]. Benabdellah dkk. [34] menggunakan nilai Dunn dan Connectivity untuk menentukan kandidat algoritme. Sulc dkk. [35] menggunakan nilai Silhouette dan Dunn untuk mengevaluasi hasil kluster. Šikonja [36] menggunakan Gower sebagai metode jarak dan nilai Silhouette sebagai metode evaluasi.

Indeks Silhouette menghitung rata-rata nilai setiap titik pada himpunan data [37]. Perhitungan nilai setiap titik adalah selisih nilai *separation* dan *compactness* yang dibagi dengan maksimum antara keduanya. Jumlah kluster yang terbaik ditunjukkan dengan nilai Silhouette yang semakin mendekati 1. Perhitungan Silhouette dinyatakan dalam Persamaan 3.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Indeks Dunn menghitung nilai minimum dari perbandingan antara nilai fungsi ketidaksamaan antara dua kluster sebagai *separation* dan nilai maksimum dari diameter kluster sebagai *compactness* [37]. Jumlah kluster terbaik ditunjukkan dengan semakin besar nilai *Dunn*. Perhitungan *Dunn* dinyatakan dalam Persamaan 4.

$$D = \frac{\min_{0 \leq i \leq n_c, 0 \leq j < n_c, i \neq j} (d(C_i, C_j))}{\max_{0 \leq k < n_c} (\text{diam}(C_k))} \quad (4)$$

Connectivity berkisar dari nol hingga tak terhingga [38]. Nilai yang lebih kecil menunjukkan kluster

kualitas yang lebih baik. Perhitungan *Connectivity* dinyatakan dalam Persamaan 5.

$$\text{conn} = \sum_{i=1}^N \sum_{j=1}^L x_{i,m_{i,j}} \quad (5)$$

Evaluasi dilakukan setelah proses mining yang dilakukan dengan algoritme k-medoids dengan dua metode jarak berbeda. Evaluasi menggunakan indeks validitas kluster dengan nilai evaluasi Silhouette, Dunn, *Connectivity* sebagai parameter.

III. HASIL DAN PEMBAHASAN

Hasil penelitian adalah berupa nilai evaluasi Silhouette, Dunn, dan *Connectivity* pada setiap metode jarak yang digunakan. Penelitian ini menggunakan 7 dataset berbeda seperti yang terdapat pada Tabel 1 dengan nilai *k* yang telah ditentukan dengan menggunakan metode *elbow* yang menunjukkan jumlah kluster terbaik, yaitu kluster 2 dan 3.

Hasil perhitungan nilai Silhouette pada metode jarak Euclidean dan Gower pada masing-masing dataset dinyatakan pada Tabel 2. Hasil evaluasi dengan nilai Silhouette menunjukkan bahwa metode jarak Euclidean jarak pada lima dataset, sedangkan metode Gower unggul pada dua dataset. Nilai tertinggi yang diperoleh Euclidean adalah 0,81, sedangkan Gower 0,85 pada dataset tenaga kesehatan. Rata-rata dari nilai Silhouette kedua metode jarak hanya berbeda 0,07 dengan jarak Euclidean lebih unggul.

Hasil perhitungan nilai *Dunn* pada metode jarak Euclidean dan Gower pada masing-masing dataset dinyatakan pada Tabel 3. Hasil evaluasi dengan nilai *Dunn* menunjukkan bahwa metode jarak Gower unggul pada lima dataset, sedangkan metode Euclidean unggul pada dua dataset. Nilai tertinggi yang diperoleh Euclidean adalah 0,58, sedangkan Gower 0,95 pada

Tabel 2. Nilai Silhouette pada metode jarak Euclidean dan Gower

No	Dataset	Euclidean	Gower	Kluster
1	Wholesale customers data	0,5453	0,4952	2
2	Hepatitis	0,5826	0,3335	2
3	Contraceptive method choice	0,5482	0,4826	2
4	Ujian nasional SMP Indonesia	0,4930	0,3337	3
5	Penyakit diare Kabupaten Karawang	0,7365	0,6360	2
6	Penyakit campak perbulan provinsi	0,5198	0,5917	3
7	Tenaga kesehatan dokter per provinsi	0,8121	0,8513	2
	Rata-rata Silhouette	0,6053	0,5320	

Tabel 3. Nilai Dunn pada metode jarak Euclidean dan Gower

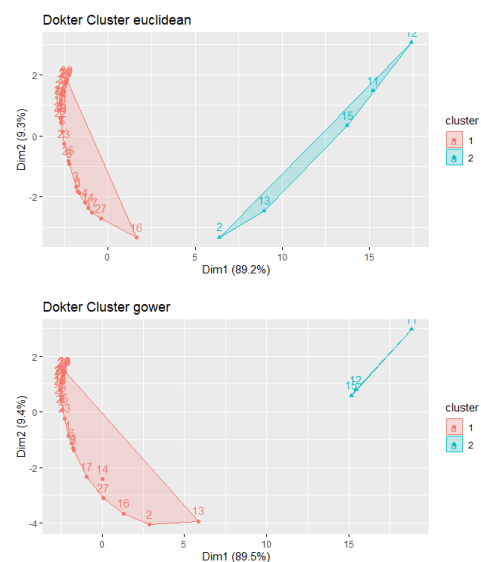
No	Dataset	Euclidean	Gower	Kluster
1	Wholesale customers data	0,0070	0,0095	2
2	Hepatitis	0,0423	0,0906	2
3	Contraceptive method choice	0,0643	0,5943	2
4	Ujian nasional SMP Indonesia	0,1900	0,0657	3
5	Penyakit diare Kabupaten Karawang	0,5869	0,2850	2
6	Penyakit campak perbulan provinsi	0,1134	0,3935	3
7	Tenaga kesehatan dokter per provinsi	0,3902	0,9791	2
	Rata-rata Dunn	0,1991	0,3453	

Tabel 4. Nilai Connectivity pada metode jarak Euclidean dan Gower

No	Dataset	Euclidean	Gower	Kluster
1	Wholesale customers data	28,2782	31,6024	2
2	Hepatitis	4,4770	40,8730	2
3	Contraceptive method choice	0,0000	0,0000	2
4	Ujian nasional SMP Indonesia	9,1810	12,8198	3
5	Penyakit diare Kabupaten Karawang	4,0861	5,9905	2
6	Penyakit campak perbulan provinsi	7,4560	5,0008	3
7	Tenaga kesehatan dokter per provinsi	5,3520	4,2869	2
	Rata-rata Connectivity	98,051	167,622	

dataset tenaga kesehatan. Rata-rata dari nilai Dunn kedua metode jarak hanya mendapatkan selisih 0,14. Hal ini menunjukkan fungsi ketidaksamaan antara dua kluster sebagai pemisah dan nilai maksimum dari diameter kluster sebagai kepadatan menunjukkan Gower lebih unggul dari Euclidean. Hasil kluster pada kedua metode jarak disajikan pada Gambar 3. Hasil ini menunjukkan semakin tinggi nilai Dunn, maka antara kluster 1 dan 2 akan saling menjauh.

Hasil perhitungan nilai Connectivity pada metode jarak Euclidean dan Gower pada masing-masing dataset dinyatakan pada Tabel 4. Hasil evaluasi dengan nilai Connectivity menunjukkan bahwa metode jarak Euclidean unggul pada lima dataset, sedangkan metode Gower unggul pada tiga dataset. Nilai terendah yang diperoleh Euclidean dan Gower adalah 0 pada dataset *contraceptive method choice*. Rata-rata dari nilai Connectivity kedua metode jarak hanya mendapatkan selisih 69,571. Hal ini menunjukkan Euclidean lebih unggul dari Gower dimana nilai terendah pada rata-rata Connectivity Euclidean adalah 98,051, sedangkan Gower 167,622.



Gambar 3. Hasil kluster dataset tenaga kesehatan

Penelitian ini sesuai dengan [9] dimana pemilihan metode jarak yang berbeda juga memberikan hasil yang berbeda. Metode jarak Gower juga dapat digunakan dengan k-medoids sebagaimana pernyataan [23]. Selain itu, metode jarak Gower juga dapat digunakan sebagai perhitungan numerik sebagaimana pernyataan [25], [26]. K-medoids juga dapat digunakan dengan berbagai pengukuran jarak sebagaimana pendapat [10]. Dibanding Gower, metode jarak Euclidean menghasilkan hasil yang baik sebagaimana dalam [16]–[19].

IV. KESIMPULAN

Metode pengukuran Euclidean lebih unggul dari Gower pada dataset bertipe numerik menggunakan algoritme klastering k-medoids. Metode pengukuran jarak Euclidean menghasilkan nilai evaluasi terbaik pada evaluasi nilai Silhouette dan Connectivity yang menunjukkan bahwa hasil klaster dengan Euclidean memiliki struktur pengelompokan data yang baik. Metode jarak Gower menjadi perhitungan jarak terendah berdasarkan nilai Silhouette dan Connectivity. Metode Gower hanya unggul pada nilai Dunn dari Euclidean pada tujuh dataset bertipe numerik yang digunakan pada penelitian ini yang menunjukkan bahwa Gower memiliki pemisahan antar klaster yang baik.

DAFTAR PUSTAKA

- [1] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan algoritma k-means dan k-medoids untuk pengelompokan data transaksi bongkar muat di provinsi Riau," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119-125, 2019. doi: [10.24014/rmsi.v5i1.7381](https://doi.org/10.24014/rmsi.v5i1.7381)
- [2] M. Anggara, H. Sujiani, and H. Nasution, "Pemilihan distance measure pada k-means clustering untuk pengelompokan member di alvaro fitness," *Jurnal Sistem dan Teknologi Informasi*, vol. 1, no. 1, pp. 1-6, 2016.
- [3] D. F. Pramesti, M. T. Furqon, and C. Dewi, "Implementasi metode k-medoids clustering untuk pengelompokan data potensi kebakaran hutan / lahan berdasarkan persebaran titik panas (hotspot)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 9, pp. 723-732, 2017.
- [4] D. Marlina, N. F. Putri, A. Fernando, and A. Ramadhan, "Implementasi algoritma k-medoids dan k-means untuk pengelompokan wilayah sebaran cacat pada anak," *Jurnal Coreit*, vol. 4, no. 2, pp. 64-71, 2018. doi: [10.24014/coreit.v4i2.4498](https://doi.org/10.24014/coreit.v4i2.4498)
- [5] F. R. Senduk, I. Indwiarti, and F. Nhita, "Clustering of earthquake prone areas in Indonesia using k-medoids algorithm," *Indonesian Journal of Computing*, vol. 4, no. 3, pp. 65-76, 2019.
- [6] R. D. Ramadhani and D. A. Januarita, "Evaluasi k-means dan k-medoids pada dataset kecil," in *Seminar Nasional Informatika dan Aplikasinya*, Bandung, Indonesia, Sept. 2019, pp. 20-24, 2017.
- [7] W. Budiaji and F. Leisch, "Simple k-medoids partitioning algorithm for mixed variable data," *Algorithms*, vol. 12, no. 117, pp. 1-15, 2019. doi: [10.3390/a12090177](https://doi.org/10.3390/a12090177)
- [8] S. Pandit and S. Gupta, "A comparative study on distance measuring approaches for clustering," *International Journal of Research in Computer Science*, vol. 2, no. 1, pp. 29-31, 2011.
- [9] S. Dahal, "Effect of different distance measures in result of cluster analysis," *Master thesis*, Aalto University School of Engineering, Finland, 2015.
- [10] M. Mohibullah, M. Z. Hossain, and M. Hasan, "Comparison of Euclidean distance function and manhattan distance function using k-medoids," *International Journal of Computer Science and Information Security*, vol. 13, no. 10, pp. 61-71, 2015.
- [11] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi k-means clustering ujian nasional sekolah menengah pertama di Indonesia Tahun 2018/2019," *Jurnal Media Informatika Budidarma*, vol. 4, no. 1, p. 51, 2020. doi: [10.30865/mib.v4i1.1784](https://doi.org/10.30865/mib.v4i1.1784)
- [12] W. Gautama, "Analisis pengaruh penggunaan manhattan distance pada algoritma clustering isodata (self-organizing data analysis technique) untuk sistem deteksi anomali trafik," *Skripsi*, Telkom University, Indonesia, 2015.
- [13] Z. Mustofa and I. S. Suasana, "Algoritma clustering k-medoids pada e-government bidang information and communication technology dalam penentuan status edgi," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 1-10, 2018.
- [14] U. Rani and S. Sahu, "Comparison of clustering techniques for measuring similarity in articles," in *3rd International Conference on Computational Intelligence & Communication Technology*, Ghaziabad, India, Feb. 2017, pp. 1711-1718. doi: [10.1109/CICT.2017.7977377](https://doi.org/10.1109/CICT.2017.7977377)
- [15] B. Ali and Y. Massmoudi, "K-means clustering based on gower similarity coefficient: A comparative study," in *5th International Conference on Modeling, Simulation and Applied Optimization*, Hammamet, Tunisia, Apr. 2013, pp. 1-5. doi: [10.1109/ICMSAO.2013.6552669](https://doi.org/10.1109/ICMSAO.2013.6552669)
- [16] M. Nishom, "Perbandingan akurasi Euclidean distance, minkowski distance, dan manhattan distance pada algoritma k-means clustering berbasis chi-square," *Jurnal Informatika*, vol. 4, no. 1, pp. 20-24, 2019. doi: [10.30591/jpit.v4i1.1253](https://doi.org/10.30591/jpit.v4i1.1253)
- [17] D. Sinwar and R. Kaushik, "Study of Euclidean and manhattan distance metrics using simple k-means clustering," *International Journal for Research in Applied Science and Engineering Technology*, vol. 2, no. 5, pp. 270-274, 2014.
- [18] A. S. Sunge, Y. Heryadi, Y. Religia, and L. Lukas, "Comparison of distance function to performance of k-medoids algorithm for clustering," in *International Conference on Smart Technology and Applications*, Surabaya, Indonesia, Feb. 2020, pp. 1-6. doi: [10.1109/ICoSTA48221.2020.1570615793](https://doi.org/10.1109/ICoSTA48221.2020.1570615793)

- [19] R. I. Fajriah, H. Sutisna, and B. K. Simpony, "Perbandingan distance space manhattan dengan euclidean pada k-means clustering dalam menentukan promosi," *Indonesian Journal on Computer and Information Technology*, vol. 4, no. 1, pp. 36-49, 2019.
- [20] S. Godara, R. Singh, and S. Kumar, "Proposed density based clustering with weighted Euclidean distance," *International Journals of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 6, pp. 409-412, 2017. doi: [10.23956/ijarcsse/V7I6/0190](https://doi.org/10.23956/ijarcsse/V7I6/0190)
- [21] Z. Šulc, J. Procházka, and M. Matějka, "Modifications of the gower similarity coefficient," in *Applications of Mathematics and Statistics in Economics*, Banská Štiavnica, Slovakia, Sept. 2016, pp. 369-377.
- [22] Z. Anna, "Acceleration of k-means clustering by dijkstra method for graph partitioning," *Thesis*, School of Information Science Nara Institute Science and Technology, Japan, 2015.
- [23] J. van den Hoven, "Clustering with optimised weights for Gower's metric," *Thesis*, University Amsterdam, Netherlands, 2016.
- [24] K. H. Jung et al., "Cluster analysis of child homicide in South Korea," *Child Abuse & Neglect*, vol. 101, 104322, 2020. doi: [10.1016/j.chiabu.2019.104322](https://doi.org/10.1016/j.chiabu.2019.104322)
- [25] A. N. Sadovski, "Detection of similar homoclimates by numerical analysis," *Bulgarian Journal of Soil Science*, vol. 4, no. 1, pp. 69-75, 2019.
- [26] A. Nowak-Brzezinska and T. Rybotycki, "Comparison of similarity measures in context of rules clustering," in *IEEE International Conference on INnovations in Intelligent SysTems and Applications*, Gdynia, Poland, Jul. 2017, pp. 235-240. doi: [10.1109/INISTA.2017.8001163](https://doi.org/10.1109/INISTA.2017.8001163)
- [27] N. N. Mohammed and A. M. Abdulazez, "Evaluation of partitioning around medoids algorithm with various distances on microarray data," in *IEEE International Conference on Internet of Things (iThings)*, Exeter, UK, Jun. 2007, pp. 1011-1016. doi: [10.1109/iThings-GreenCom-CPSCoM-SmartData.2017.155](https://doi.org/10.1109/iThings-GreenCom-CPSCoM-SmartData.2017.155)
- [28] C. W. Putra and R. Rian, "Implementasi data mining pemilihan pelanggan potensial menggunakan algoritma k-means," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 1, no. 1, pp. 72-77, 2018. doi: [10.31539/intecom.v1i1.141](https://doi.org/10.31539/intecom.v1i1.141)
- [29] F. L. Sibuea and A. Sapta, "Pemetaan siswa berprestasi menggunakan metode k-means clustering," *JURTEKSI*, vol. 4, no. 1, pp. 85-92, 2017. doi: [10.33330/jurteksi.v4i1.28](https://doi.org/10.33330/jurteksi.v4i1.28)
- [30] R. Fitriani and N. Rosmawanti, "Penerapan algoritma euclidean distance untuk pemilihan paket internet berdasarkan wilayah," *Progresif*, vol. 13, no. 1, pp. 1651-1662, 2017.
- [31] A. Skabar, "Clustering mixed-attribute data using random walk," *Procedia Computer Science*, vol. 108, pp. 988-997, 2017. doi: [10.1016/j.procs.2017.05.083](https://doi.org/10.1016/j.procs.2017.05.083)
- [32] N. Putu, E. Merliana, and A. J. Santoso, "Analisa penentuan jumlah cluster terbaik pada metode k-means," in *Seminar Nasional Multi Disiplin Ilmu*, Semarang, Indonesia, Aug. 2015, pp. 978-979.
- [33] A. D. Savitri, F. A. Bachtar, and N. Y. Setiawan, "Segmentasi pelanggan menggunakan metode k-means clustering berdasarkan model rfm pada klinik kecantikan (studi kasus: Belle Crown Malang)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, pp. 2957-2966, 2018.
- [34] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia Computer Science*, vol. 148, pp. 291-302, 2019. doi: [10.1016/j.procs.2019.01.022](https://doi.org/10.1016/j.procs.2019.01.022)
- [35] Z. Šulc, M. Matějka, J. Procházka, and H. Řezanková, "Evaluation of the Gower coefficient modifications in hierarchical clustering," *Metodološki Zvezki*, vol. 14, no. 1, pp. 37-48, 2017.
- [36] M. R. Šikonja, "Dataset comparison workflows," *International Journal of Data Science*, vol. 3, no. 2, p. 126, 2018. doi: [10.1504/IJDS.2018.10013385](https://doi.org/10.1504/IJDS.2018.10013385)
- [37] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, "Kajian indeks validitas pada algoritma k-means enhanced dan k-means MMCA," *Prosiding Seminar Nasional Matematika*, vol. 2, pp. 161-170, 2019.
- [38] S. M. Kim, M. I. Peña, M. Moll, G. Giannakopoulos, G. N. Bennett, and L. E. Kavraki, "An evaluation of different clustering methods and distance measures used for grouping metabolic pathways," in *International Conference on Bioinformatics and Computational Biology*, Kuala Lumpur, Malaysia, Feb. 2016, pp. 115-122.