JISKA (Jurnal Informatika Sunan Kalijaga)
Vol. 11, No. 1, JANUARI, 2026, Pp. 32 – 43
ISSN: 2527 – 5836 (print) | 2528 – 0074 (online)

Komparasi *Distance Measure* pada K-Means dalam Klasterisasi Peserta KB Aktif

Mochammad Anshori [(1)*], Afifah Vera Ferencia Fitria Ningrum [(2)], Risqy Siwi Pradini [(3)]
Departemen Informatika, Institut Teknologi, Sains, dan Kesehatan RS. Dr. Soepraoen Kesdam V/BRW, Malang, Indonesia
e-mail : {moanshori,risqypradini}@itsk-soepraoen.ac.id, afifahveraferencia@gmail.com.
* Penulis korespondensi.
Artikel ini diajukan 3 Februari 2025, direvisi 12 November 2025, diterima 15 November 2025, dan dipublikasikan 25 Januari 2026.

Abstract

The rapid population growth in Indonesia poses significant challenges to public welfare, economic stability, and sustainable development. The Family Planning program aims to regulate population growth through various contraceptive methods; however, participation rates often differ across regions. Understanding these variations is crucial for designing targeted interventions. This study investigates how different distance measures in the K-Means clustering algorithm affect the segmentation quality of KB participants in Kalirejo Village, Lawang District. Eight distance metrics—Euclidean, Manhattan, Minkowski, Chebyshev, Mahalanobis, Bray-Curtis, Canberra, and Cosine—were compared using standardized data from the local BKKBN office (January–September). Cluster validity was evaluated using the Silhouette Coefficient across k=2–10. Results show that the Manhattan distance with k=2 achieved the best clustering quality (SC = 0.7191), effectively distinguishing participant groups by contraceptive method preference. The study highlights the importance of selecting suitable distance measures to improve data-driven policy and decision-making in family planning management.

Keywords: K-Means, Clustering, Silhouette Coefficient, Distance Measure, Manhattan

Abstrak

Pertumbuhan penduduk yang pesat di Indonesia menimbulkan tantangan besar terhadap kesejahteraan masyarakat, stabilitas ekonomi, dan pembangunan berkelanjutan. Program Keluarga Berencana (KB) bertujuan mengendalikan laju pertumbuhan penduduk melalui berbagai metode kontrasepsi, namun tingkat partisipasi masyarakat sering kali berbeda antarwilayah. Pemahaman terhadap variasi tersebut sangat penting untuk merancang intervensi yang lebih tepat sasaran. Penelitian ini mengkaji pengaruh penggunaan berbagai ukuran jarak (*distance measure*) pada algoritma K-Means terhadap kualitas segmentasi peserta KB di Desa Kalirejo, Kecamatan Lawang. Delapan jenis ukuran jarak—Euclidean, Manhattan, Minkowski, Chebyshev, Mahalanobis, Bray-Curtis, Canberra, dan Cosine—dibandingkan menggunakan data terstandarisasi dari BKKBN setempat (Januari–September). Validitas klaster dievaluasi dengan Silhouette Coefficient pada rentang k=2–10. Hasil menunjukkan bahwa ukuran jarak Manhattan dengan k=2 menghasilkan kualitas klaster terbaik (SC = 0,7191), yang secara efektif membedakan kelompok peserta berdasarkan preferensi metode kontrasepsi. Penelitian ini menegaskan pentingnya pemilihan ukuran jarak yang tepat untuk meningkatkan kualitas analisis berbasis data serta mendukung pengambilan keputusan kebijakan dalam pengelolaan program KB.

Kata Kunci: K-Means, *Clustering*, Silhouette Coefficient, *Distance Measure*, Manhattan

1. PENDAHULUAN

Salah satu program pemerintah yang bertujuan untuk mengontrol pertumbuhan penduduk dan meningkatkan kesejahteraan masyarakat adalah Keluarga Berencana (KB). Program ini mengatur jarak kelahiran untuk mengimbangi jumlah penduduk dan ketersediaan sumber daya. (Munawar et al., 2024; Sumarsih, 2023). Dengan mengendalikan angka kelahiran, Program KB diharapkan dapat menghentikan peningkatan populasi yang berlebihan dan mencegah dampak negatif yang dapat muncul akibat lonjakan populasi yang tidak terkendali. Selain itu, program KB

meningkatkan kualitas hidup keluarga dan individu dengan memberikan akses ke perawatan kesehatan reproduksi berkualitas tinggi (Tifannii et al., 2020).

Pertumbuhan penduduk yang tidak terkendali dalam konteks sosial dan ekonomi dapat menyebabkan berbagai masalah, seperti peningkatan kriminalitas, kesenjangan kesejahteraan dan makin terbatasnya lapangan kerja (Maharani & Yotenka, 2024). Oleh karena itu pemerintah mencanangkan program KB sebagai langkah strategis untuk pengelolaan kuantitas penduduk. Sebagai lembaga yang bertanggung jawab atas pelaksanaan program keluarga berencana di Indonesia, Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) berusaha untuk meningkatkan kinerja program melalui penggunaan berbagai teknik berbasis data dan teknologi (Lino et al., 2021). Untuk mendapatkan pemahaman tentang pola partisipasi dan efektivitas program di berbagai wilayah, peserta KB diklasifikasikan atau dibagi menjadi kelompok.

Desa Kalirejo di Kecamatan Lawang merupakan salah satu wilayah yang menerapkan program KB yang diinisiasi oleh Puskesmas setempat. Puskesmas ini berperan sebagai Klinik Keluarga Berencana (KKB) yang menyediakan layanan kontrasepsi bagi masyarakat setempat. Meskipun program ini telah berjalan, masih terdapat kecenderungan tingkat kelahiran yang tinggi di beberapa bagian desa. Hal ini menunjukkan adanya kebutuhan untuk memahami distribusi dan pola partisipasi peserta KB guna meningkatkan efektivitas penyuluhan dan layanan KB yang diberikan. Klasterisasi peserta KB aktif adalah langkah penting dalam mengoptimalkan program ini. Ini memungkinkan untuk menjangkau kelompok sasaran dengan lebih akurat.

Sebelum ini, penelitian klasterisasi peserta KB aktif di Desa Kalirejo telah dilakukan dengan menggunakan metode K-Means dengan Euclidean distance sebagai ukuran jarak. Hasil penelitian tersebut menunjukkan bahwa *silhouette coefficient* yang diperoleh sebesar 0,447 dengan jumlah klaster optimal sebanyak dua (Ningrum et al., 2025). Nilai *silhouette coefficient* ini berada dalam rentang -1 hingga 1, di mana semakin mendekati angka 1 menunjukkan bahwa klaster yang terbentuk semakin baik. Di sisi lain, hasil masih belum ideal, jadi diperlukan penelitian lebih lanjut untuk meningkatkan keakuratan dan keefektifan model klasterisasi.

Dalam metode K-Means, pemilihan ukuran jarak atau *distance measure* memiliki peran penting dalam menentukan kualitas klaster yang terbentuk. Oleh karena itu, diperlukan eksplorasi lebih lanjut terhadap berbagai *distance measure* yang dapat digunakan selain Euclidean distance. Beberapa *distance measure* yang umum digunakan dalam klasterisasi mencakup *Manhattan, Minkowski, Chebyshev, Mahalanobis, Bray-Curtis, Canberra dan Cosine distance* (Argiento et al., 2025; Idrus et al., 2022; Jollyta et al., 2023; Kumala & Rahning Putri, 2023; Pangestu & Fitriani, 2022; Shapcott, 2024; Telsiz Kayaoğlu & Eroğlu, 2024; Wurdianarto et al., 2014). Masing-masing metode memiliki karakteristik yang berbeda dan dapat memberikan hasil yang lebih baik pada kondisi data tertentu.

Penelitian terdahulu menunjukkan bahwa variasi *distance measure* dapat memberikan dampak yang signifikan terhadap hasil klasterisasi. Studi sebelumnya membandingkan perhitungan *euclidean*, *manhattan, dan cosine distance* dalam pengelompokan data dan menemukan bahwa *manhattan distance* memiliki performa yang lebih baik pada data dengan distribusi yang tidak merata (Pangestu & Fitriani, 2022). Riset lainnya telah melakukan perbandingan antara *chebyshev* dengan *manhattan* dan menghasilkan *chebyshev* menghasilkan klaster yang lebih sempurna (Solikhun et al., 2025).

Meskipun banyak penelitian telah mengeksplorasi perbandingan *distance measure* dalam klasterisasi, masih terdapat kesenjangan penelitian dalam penerapan metode ini pada klasterisasi peserta KB. Kebanyakan penelitian sebelumnya lebih fokus pada aplikasi *distance measure* dalam bidang kesehatan secara umum atau data non-klinis, Untuk meningkatkan efektivitas klasterisasi peserta KB aktif di Desa Kalirejo, penelitian ini bertujuan untuk mempelajari dan membandingkan berbagai ukuran jarak dalam metode K-Means.

Penelitian ini bertujuan untuk mengelompokkan peserta KB aktif di Desa Kalirejo dengan menerapkan metode K-Means dan berbagai ukuran jarak (*distance measure*) guna menentukan metode mana yang menghasilkan performa terbaik. Studi ini memberikan kontribusi dalam memperluas pemahaman mengenai penggunaan ukuran jarak dalam proses pengelompokan data, sekaligus mendukung pengambilan keputusan yang berbasis data dalam program KB. Diharapkan, temuan penelitian ini dapat memberikan masukan berharga dalam perencanaan dan pelaksanaan program KB yang lebih terarah, serta meningkatkan efektivitas layanan KB di tingkat desa.

2. METODE PENELITIAN

Penelitian ini melibatkan serangkaian langkah yang penting untuk mempersiapkan dan merencanakan studi secara menyeluruh. Penelitian ini dirancang untuk mengatasi permasalahan yang diidentifikasi dalam studi. Secara umum, terdapat lima tahapan utama yang dilakukan dalam penelitian ini, seperti yang diilustrasikan pada Gambar 1. Tahapan-tahapan tersebut meliputi pengumpulan data, praproses data, dan pembuatan model klasterisasi menggunakan metode K-Means dengan berbagai macam *distance measure* dan terakhir adalah evaluasi dengan menggunakan s*ilhouette coefficient*.

Gambar 1 Metode Penelitian

2.1 Pengumpulan Data

Penelitian ini memanfaatkan data sekunder yang bersumber dari Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) di Desa Kalirejo, Kecamatan Lawang. Data yang dikumpulkan meliputi informasi tentang lokasi tempat tinggal peserta KB (RT dan RW) serta jenis alat kontrasepsi yang digunakan, seperti IUD, MOW, MOP, implant, suntik, pil, dan kondom (Maulana, 2020). Data ini dikumpulkan dalam rentang waktu Januari hingga September. Untuk menjaga privasi individu, penelitian ini tidak mengikutsertakan data personal warga. Selanjutnya data dilakukan proses pembersihan.

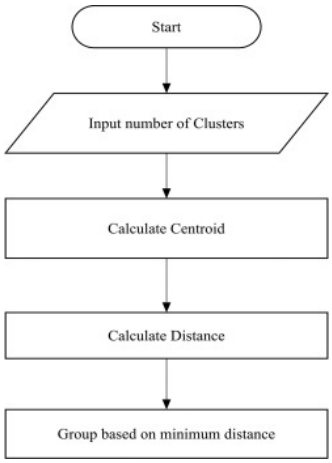
2.2 Praproses Data

Algoritme K-Means sangat sensitif terhadap skala data (Shapcott, 2024), sehingga tahap praproses dilakukan untuk memastikan bahwa data memiliki distribusi yang lebih seragam. Salah satu teknik yang digunakan adalah standardisasi data dengan metode z-score normalization, yang dapat mengurangi *skewness* dalam distribusi data dan membuat setiap atribut memiliki rata-rata nol serta varians satu (Ghosh, 2022). Proses ini bertujuan untuk menghindari bias akibat perbedaan skala antar atribut. Formula dari standardisasi ditunjukkan pada Pers (1), di mana x adalah data awal, μ adalah rerata dari atribut, σ adalah standar deviasi dan x' adalah data baru hasil standardisasi z-score.

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

2.3 Klasterisasi K-Means

Klasterisasi merupakan teknik yang digunakan untuk membagi data ke dalam beberapa kelompok berdasarkan karakteristik tertentu (Andriyani et al., 2024) dan termasuk metode yang prominent pada bidang ilmu komputer (Wala et al., 2024). Algoritma K-Means bekerja dengan mengelompokkan data ke dalam sejumlah klaster, di mana data dalam satu klaster memiliki kesamaan (homogenitas) dalam hal karakteristik (J. A. Muttaqin et al., 2023; W. W. W. Muttaqin et al., 2023). Metode ini dikenal efisien karena kecepatannya dalam memproses data dan kemampuannya untuk menangani data dalam skala besar. Namun, metode ini memiliki keterbatasan dalam hal akurasi ketika menghadapi noise atau data yang terisolasi (Hutagalung, 2022; Hutagalung & Sonata, 2021; Pratistha & Kristianto, 2024).



Gambar 2 Flowchart Algoritma K-Means

Cara kerja pengelompokan K-Means cukup sederhana. Gambar 2 menunjukkan mekanisme K-Means. Mengikuti gambar tersebut, pertama-tama harus menentukan k klaster. Kemudian menghitung centroid dengan inisialisasi acak. Langkah ketiga adalah menghitung jarak antara titik data ke centroid menggunakan ukuran jarak yang dipilih, jarak *euclidean* umumnya digunakan. Langkah terakhir adalah mengelompokkan klaster berdasarkan angka rerata terdekat (Rawal et al., 2021). Proses ini akan berulang hingga tidak ada pembaruan yang ditemukan untuk centroid.

2.4 *Distance Measure*

Berbagai macam pengukuran jarak yang akan digunakan dalam percobaan ini adalah *Euclidean distance, Manhattan distance, Minkowski distance, Chebyshev distance, Mahalanobis distance, Bray-Curtis distance, Canberra distance,* dan *Cosine distance*. Euclidean distance merupakan jarak yang paling umum digunakan dan paling sederhana antara titik, terutama untuk variabel kontinu. Rumus perhitungan Euclidean distance mengacu pada Pers (2). Canberra distance adalah fungsi metrik yang sering digunakan untuk distribusi di sekitar daerah asal, di mana perhitungannya dilakukan dengan membagi perbedaan absolut antara variabel dari dua objek dengan jumlah nilai variabel absolut tersebut. Rumus Canberra distance ditampilkan pada Pers (3). Manhattan distance, yang juga disebut sebagai *City-Block distance*, adalah pengukuran absolut yang dihasilkan berdasarkan jumlah selisih antara dua titik data, dengan rumus yang ditunjukkan pada Pers (4).

$$d_{euclidean}(i,j) = \sqrt{\sum_{k=1}(x_{ik} - x_{jk})^2} \quad (2)$$

$$d_{canberra}(i,j) = \sum_{k=1} \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|} \quad (3)$$

$$d_{manhattan}(i,j) = \sum_{k=1} |x_{ik} - x_{jk}| \quad (4)$$

Selanjutnya, Cosine distance digunakan untuk menghitung jarak kosinus antara dua data point dan biasa dikenal dengan *cosine similarity* dengan formula yang ditunjukkan pada Pers (5). Chebyshev distance, atau jarak metrik maksimum, menghitung perbedaan absolut maksimum antara koordinat dua titik di sepanjang dimensi apa pun. Metode ini tangguh terhadap outlier dan dapat menangani data dengan skala yang bervariasi, sebagaimana ditunjukkan pada Pers (6). Bray-Curtis distance pada dasarnya menghitung perbedaan antara dua sampel berdasarkan komposisi spesiesnya, dengan nilai berkisar dari 0 (benar-benar mirip) hingga 1 (benar-benar berbeda). Formula Bray-Curtis ditunjukkan pada Pers (7).

$$d_{cosine}(i,j) = 1 - \frac{\sum_{k=1}(x_{ik} - x_{jk})}{\sqrt{\sum_{k=1} {x_{ik}}^2}\sqrt{\sum_{k=1} {x_{jk}}^2}} \quad (5)$$

$$d_{chebyshev}(i,j) = \max_k |x_{ik} - x_{jk}| \quad (6)$$

$$d_{bray-curtis}(i,j) = \frac{\sum_{k=1} |x_{ik} - x_{jk}|}{\sum_{k=1}(x_{ik} + x_{jk})} \quad (7)$$

Selain itu, Minkowski distance adalah pengukuran jarak dengan generalisasi pengukuran dari Euclidean dan Manhattan. Formula dari Minkowski distance ditunjukkan pada Pers (8). Terakhir, Mahalanobis distance adalah pengukuran yang digunakan untuk mengukur jarak antara data point dengan distribusinya dengan memanfaatkan *covariance* data, membuatnya cocok untuk karakteristik data dengan varians yang berbeda. Formula dari Mahalanobis distance ditunjukkan pada Pers (9).

$$d_{minkowski}(i,j) = (\sum_{k=1}(x_{ik} - x_{jk})^q)^{1/q} \quad (8)$$

$$d_{mahalanobis}(i,j) = \sqrt{(x_{ik} - x_{jk})^T Cov^{-1}(x_{ik} - x_{jk})} \quad (9)$$

2.5 Evaluasi

Eksperimen yang akan dilakukan adalah dengan mencoba berbagai macam *distance measure* yang telah dibahas sebelumnya. Pengujian juga dilakukan untuk mendapatkan klaster terbaik berdasarkan parameter *k* dari K-Means. Pengujian parameter *k* yang akan dilakukan dengan rentang nilai antara 2 hingga 10. Tahap berikutnya untuk mengetahui seberapa bagus klaster dengan dilakukan evaluasi model menggunakan *silhoutte coefficient* (SC). Koefisien siluet rata-rata dihitung menggunakan jarak intra-klaster dan jarak terdekat ke klaster untuk setiap titik data (Shahapure & Nicholas, 2020).

$$\mathrm{S} = \frac{(b - a)}{\max(a,b)} \quad (10)$$

Merujuk pada Pers (10) adalah formula dari SC, di mana maksimum (a, b) adalah nilai maksimum antara a dan b, sedangkan b adalah rata-rata jarak antara sampel dan semua sampel dalam klaster yang sama. Dalam analisis penelitian ini, SC menunjukkan pengelompokan data; SC dengan nilai mendekati +1 menunjukkan bahwa titik data berada di tingkat tinggi, sedangkan SC

dengan nilai mendekati 0 menunjukkan bahwa titik data mungkin berada di klaster lain. Interpretasi nilai SC adalah sebagai berikut (Kumala & Rahning Putri, 2023):

- 0.7 0,7 ≤ SC ≤ 1: Struktur klaster yang kuat.
- 0,5 ≤ SC < 0,7: Struktur klaster yang sedang.
- 0,25 ≤ SC < 0,5: Struktur klaster yang lemah.
- SC < 0,25: Tidak ada struktur klaster yang jelas.

Setelah implementasi model dan evaluasi klasterisasi, dilakukan analisis terhadap hasil untuk menentukan metode *distance measure* yang memberikan hasil terbaik. Hasil diperbandingkan berdasarkan nilai SC untuk masing-masing *distance measure* dan jumlah klaster k yang bervariasi antara 2 hingga 10. Visualisasi klaster menggunakan grafik *silhouette plot* dan grafik sebaran klaster dibuat untuk memahami distribusi data dalam setiap metode. Dengan metodologi yang sistematis ini, penelitian ini bertujuan untuk mengidentifikasi *distance measure* yang paling sesuai dalam klasterisasi peserta KB aktif, sehingga dapat memberikan rekomendasi untuk penerapan yang lebih akurat dalam program KB berbasis data.

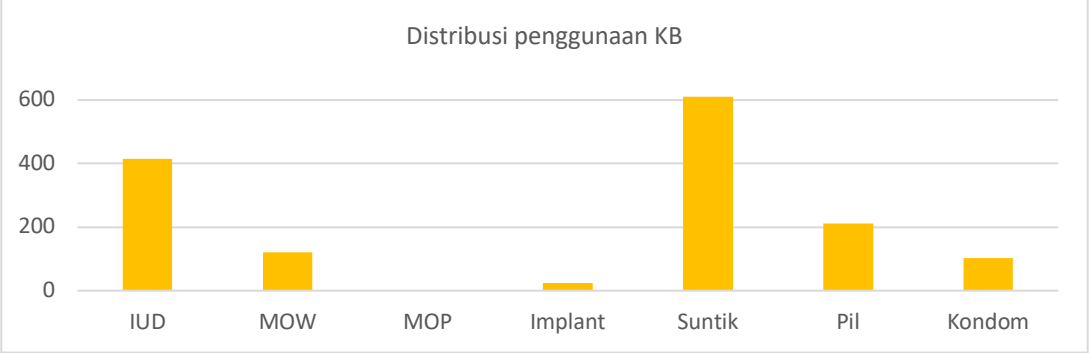
3. HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini adalah untuk mengevaluasi seberapa efektif berbagai ukuran jarak metode K-Means dalam klasterisasi peserta KB aktif di Desa Kalirejo. Hasil analisis disajikan dalam beberapa langkah. Ini termasuk penjelasan tentang dataset, hasil eksperimen klasterisasi, pengujian model menggunakan SC, dan analisis tambahan tentang pilihan jumlah klaster dan metrik jarak yang paling cocok. Dataset yang digunakan terdiri dari data peserta KB di desa Kalirejo Lawang, yang diuraikan secara rinci dalam Tabel 1.

Tabel 1 Rentang Data

Fitur	Data range	Data Type
RT	1 – 8	Numerik
RW	1 – 16	Numerik
IUD	0 – 39	Numerik
MOW	0 – 7	Numerik
MOP	0 – 1	Numerik
Implant	0 – 3	Numerik
Suntik	0 – 86	Numerik
PIL	0 – 41	Numerik
Kondom	0 – 14	Numerik

Nama fitur dan rentang nilainya tercantum dalam Tabel 1. Terdapat sembilan fitur yang diketahui: RT, RW, IUD, MOW, MOP, Implant, Suntik, PIL, dan Kondom. Rentang data untuk fitur RT adalah 1–8, sedangkan RW adalah 1–16. Pengguna alat kontrasepsi IUD berkisar antara 0 dan 39, sementara MOW berkisar antara 0 dan 7, dan MOP berkisar antara 0 dan 1. Pengguna alat kontrasepsi implant berkisar antara 0 dan 3, sedangkan pengguna suntik berkisar antara 0 dan 86. Pengguna pil berkisar antara 0 dan 41, dan pengguna kondom berkisar antara 0 dan 14. Untuk memahami seberapa besar variasi penggunaan alat kontrasepsi di setiap RT/RW, rentang ini memberikan gambaran tentang penyebaran penggunaan alat kontrasepsi di Desa Kalirejo Lawang.



Gambar 3 Pengguna Alat Kontrasepsi Desa Kalirejo Lawang

Di Desa Kalirejo Lawang, pengguna alat kontrasepsi bervariasi di RT dan RW, seperti yang ditunjukkan pada Gambar 3. Metode kontrasepsi yang paling umum digunakan (460 peserta) adalah suntikan diikuti oleh IUD (349 peserta) dan pil (171 peserta). Sementara itu, metode MOP memiliki jumlah pengguna yang paling sedikit (1 peserta), menunjukkan tingkat adopsi yang rendah untuk metode ini di kalangan peserta KB.

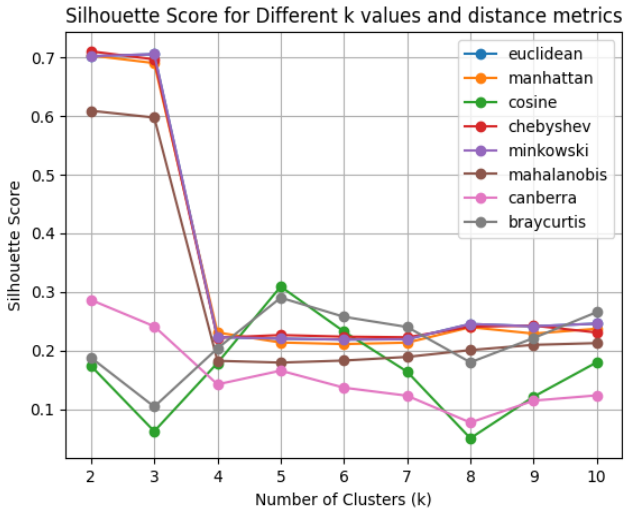
Selanjutnya eksperimen klasterisasi dengan K-Means diterapkan sesuai dengan skenario pengujian pada metodologi sebelumnya. Eksperimen klasterisasi dilakukan dengan menerapkan metode K-Means menggunakan berbagai *distance measure*, dengan jumlah klaster k yang bervariasi dari 2 hingga 10. Evaluasi kualitas klasterisasi dilakukan menggunakan SC. Tabel 2 menyajikan hasil eksperimen untuk masing-masing *distance measure*.

Tabel 2 Hasil Eksperimen Pengujian dengan Evaluasi SC

Distance	2	3	4	5	6	7	8	9	10
euclidean	**0,6990**	0,2321	0,2173	0,2172	0,2163	0,2353	0,2255	0,2415	0,2405
manhattan	**0,7191**	0,2407	0,2264	0,2270	0,2255	0,2208	0,2202	0,2488	0,2289
cosine	0,1716	**0,2166**	0,1826	0,0926	0,0387	0,1711	0,0803	0,1444	0,1843
chebyshev	**0,6821**	0,2328	0,2145	0,2085	0,2105	0,2453	0,2190	0,2202	0,2290
minkowski	**0,6990**	0,2321	0,2173	0,2172	0,2163	0,2353	0,2255	0,2415	0,2405
mahalanobis	**0,5887**	0,1913	0,1784	0,1810	0,1879	0,1949	0,2022	0,1884	0,1927
canberra	**0,2664**	0,1490	0,1457	0,1212	0,0981	0,1063	0,0989	0,1131	0,1163
braycurtis	0,1938	0,2529	0,2078	0,1748	0,1499	0,2357	0,2079	0,2532	**0,2746**

Berdasarkan Tabel 2 yang disajikan, eksperimen ini bertujuan untuk mengevaluasi performa berbagai ukuran jarak dalam algoritma K-Means dengan menggunakan *silhouette coefficient* (SC) sebagai metrik evaluasi. SC mengukur seberapa baik suatu klaster terpisah dan seberapa kohesif klaster tersebut. Semakin tinggi nilainya, semakin baik hasil klasterisasi. Dari hasil yang diperoleh, jarak Manhattan memiliki nilai SC tertinggi pada k=2 sebesar 0,7191, mengindikasikan bahwa dengan kecilnya jumlah klaster, Manhattan memberikan pemisahan terbaik. Namun, seiring bertambahnya jumlah klaster (k), performa Manhattan menurun, dan untuk k=10, nilai SC menurun menjadi 0,2289. Sebaliknya, jarak Bray-Curtis memiliki performa terbaik pada k=10 dengan nilai SC sebesar 0,2746, menunjukkan bahwa jarak ini lebih efektif untuk jumlah klaster yang lebih besar.

Sementara itu, jarak Euclidean dan Minkowski memiliki pola nilai yang sama, karena Minkowski adalah generalisasi dari Euclidean. Keduanya memiliki performa cukup baik pada k=2 (0,6990), tetapi menurun seiring bertambahnya k. Ukuran jarak Cosine memiliki performa paling buruk, dengan nilai SC yang rendah di hampir semua nilai k, terutama pada k=5 di mana SC turun drastis ke 0,0387. Hal ini menunjukkan bahwa *cosine similarity* tidak cocok untuk pengelompokan berbasis jarak dalam konteks ini. Selain itu, jarak Canberra memiliki nilai tinggi di k=2 (0,2664), tetapi turun signifikan ketika k bertambah, menunjukkan bahwa Canberra kurang stabil dalam menangani klaster yang lebih kompleks.



Gambar 4 Grafik Hasil Silhoutte Score untuk Setiap Klaster dan *Distance Measure*

Grafik SC terhadap berbagai nilai k dan metrik jarak dalam metode K-Means ditunjukkan pada Gambar 4. Grafik tersebut menunjukkan bahwa pemilihan jumlah klaster dan metrik jarak berperan krusial dalam menentukan kualitas klasterisasi. Pada k=2, terlihat bahwa metrik Euclidean, Manhattan, Chebyshev, dan Minkowski memiliki skor tertinggi, mendekati 0,7, yang mengindikasikan bahwa data terbagi dengan baik ke dalam dua klaster. Metrik Mahalanobis juga menunjukkan performa cukup baik pada k=2, meskipun sedikit lebih rendah dibandingkan metrik lainnya. Namun, ketika nilai k meningkat menjadi 3, terjadi penurunan drastis pada hampir semua metrik, terutama pada Cosine yang memiliki skor terendah di antara semua metode. Penurunan ini menunjukkan bahwa dengan lebih banyak klaster, pemisahan antar kelompok menjadi kurang jelas, sehingga menurunkan koherensi internal klaster.

Setelah k=4, skor Silhouette cenderung stabil dengan fluktuasi kecil, tetapi tetap jauh lebih rendah dibandingkan nilai awal pada k=2. Metrik Cosine mengalami sedikit peningkatan pada k=5, meskipun secara keseluruhan tetap lebih rendah dibandingkan metrik lainnya. Sementara itu, metrik Canberra secara konsisten menunjukkan skor yang lebih rendah dibandingkan metrik lainnya, mengindikasikan bahwa metode ini kurang efektif dalam menentukan pemisahan klaster pada dataset ini. Metrik Bray-Curtis menunjukkan fluktuasi yang lebih tinggi dibandingkan metrik lainnya, dengan sedikit peningkatan pada nilai k yang lebih besar, tetapi tetap dalam rentang yang relatif rendah.

Secara keseluruhan, hasil ini menunjukkan bahwa jumlah klaster yang optimal berdasarkan Silhouette Score adalah k=2, karena setelah nilai ini, kualitas klasterisasi menurun signifikan. Selain itu, metrik jarak seperti Euclidean, Manhattan, Chebyshev, dan Minkowski lebih cocok digunakan untuk dataset ini dibandingkan Cosine, Canberra, dan Bray-Curtis. Oleh karena itu, dalam pemilihan parameter untuk K-Means, perlu dilakukan analisis mendalam terhadap sifat data agar dapat menentukan kombinasi jumlah klaster dan metrik jarak yang optimal untuk hasil klasterisasi yang lebih tepat.

Tabel 3 Hasil Pengujian Terbaik Tiap Klaster dan *Distance Measure*

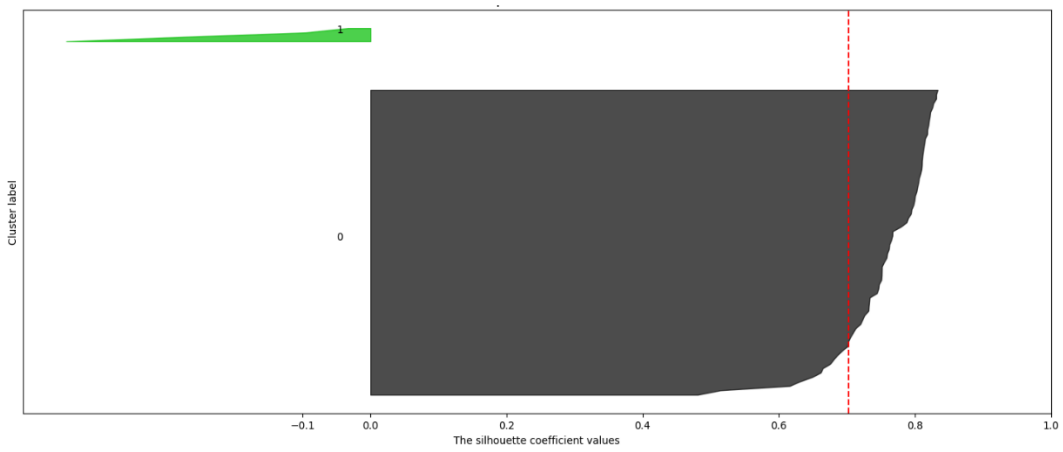
Distance Measure	Silhouette Coefficient (SC)
euclidean (k = 2)	0,6990
manhattan (k = 2)	**0,7191**
cosine (k = 3)	0,2166
chebyshev (k = 2)	0,6821
minkowski (k = 2)	0,6990

mahalanobis (k = 2)	0,5887
canberra (k = 2)	0,2664
braycurtis (k = 10)	0,2746

Berdasarkan Tabel 3, hasil klasterisasi K-Means dengan berbagai metrik jarak dan jumlah klaster k terbaik, dapat dilakukan analisis terhadap efektivitas masing-masing pendekatan dalam menghasilkan klaster yang lebih terstruktur. SC digunakan sebagai indikator kualitas klaster, dengan nilai yang lebih tinggi menunjukkan bahwa objek dalam satu klaster lebih mirip satu sama lain dibandingkan dengan objek di klaster lain. Dari tabel, terlihat bahwa metrik jarak Manhattan (k=2) memiliki nilai SC tertinggi, yaitu 0,7191, menunjukkan bahwa pendekatan ini menghasilkan klaster yang paling baik dibandingkan dengan metrik lainnya. Hal ini mengindikasikan bahwa dalam ruang vektor yang digunakan, penggunaan jarak Manhattan lebih efektif dalam mengelompokkan data dengan batasan yang lebih jelas dibandingkan metrik lainnya seperti Euclidean (0,6990) atau Minkowski (0,6990), yang memiliki karakteristik perhitungan jarak yang serupa.

Sebaliknya, metrik jarak Cosine (k=3) menghasilkan nilai SC terendah, yaitu 0,2166, yang menunjukkan bahwa metode ini kurang efektif dalam mengelompokkan data dalam konteks yang digunakan. Hal ini bisa terjadi karena jarak Cosine lebih berfokus pada sudut antara vektor daripada jauh jarak, sehingga mungkin kurang sesuai untuk struktur data yang berbasis perbedaan absolut antar titik. Jarak Chebyshev (k=2) juga menunjukkan performa yang cukup baik dengan SC 0,6821, yang berarti pendekatan ini dapat menangkap pola tertentu dalam data dengan cukup efektif. Namun, metrik lain seperti Mahalanobis (k=2) menghasilkan nilai yang lebih rendah (0,5887), yang mungkin disebabkan oleh sensitivitasnya terhadap korelasi antar variabel dan distribusi data yang tidak sesuai dengan asumsi Mahalanobis. Dua metrik lainnya, Canberra (k=2) dan Bray-Curtis (k=10), menunjukkan hasil yang lebih rendah, masing-masing dengan nilai 0,2664 dan 0,2746. Hal ini menunjukkan bahwa pendekatan berbasis perbandingan relatif atau distribusi antar nilai mungkin kurang optimal dalam menangkap pola klasterisasi dalam dataset yang digunakan.

Secara keseluruhan, analisis ini menunjukkan bahwa pemilihan metrik jarak yang tepat dalam K-Means sangat berpengaruh terhadap kualitas klasterisasi. Berdasarkan evaluasi yang dilakukan, jarak Manhattan dengan k=2 terbukti sebagai pilihan terbaik dalam konteks ini. Temuan ini menunjukkan bahwa data dalam dataset tersebut lebih sesuai untuk diukur dengan perbedaan absolut dibandingkan dengan pendekatan berbasis sudut atau distribusi relatif.

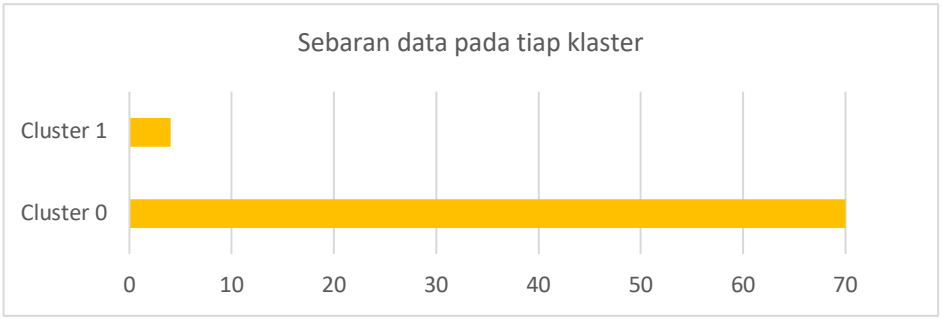


Gambar 5 Hasil Silhoutte Plot

Gambar 5 menunjukkan grafik *sillhouette plot* untuk menampilkan hasil klasterisasi menggunakan K-Means dengan k=2 dan metrik jarak Manhattan, yang sebelumnya diketahui memberikan nilai SC tertinggi (0,7191) dibandingkan metrik lainnya. Grafik ini membantu dalam memahami

bagaimana setiap sampel dalam klaster terdistribusi berdasarkan koefisien silhouette, yang mencerminkan seberapa baik objek berada dalam klasternya dibandingkan dengan klaster lain. Dari grafik, dapat diamati bahwa terdapat dua klaster utama, yaitu klaster 0 (hitam) dan klaster 1 (hijau). Klaster 0 mendominasi, dengan sebagian besar sampelnya memiliki nilai SC yang bervariasi, namun tetap berada dalam rentang yang menunjukkan kualitas klasterisasi yang cukup baik. Sementara itu, klaster 1 memiliki jumlah sampel yang jauh lebih sedikit tetapi memiliki nilai silhouette yang tinggi dan stabil, menandakan bahwa objek dalam klaster ini lebih jelas terpisah dari klaster lain.

Garis merah putus-putus yang ditampilkan pada sekitar 0,7191 menunjukkan rata-rata SC, yang relatif tinggi. Hal ini mengindikasikan bahwa mayoritas titik data memiliki pemisahan yang baik antara klaster, dengan hanya sebagian kecil sampel yang memiliki nilai SC negatif atau mendekati nol. Beberapa sampel dalam klaster 0 memiliki nilai SC yang mendekati nol atau negatif, yang berarti ada kemungkinan objek-objek tersebut berada di batas antara klaster dan memiliki kemiripan dengan klaster lainnya. Dari perspektif analisis klasterisasi, hasil ini mengonfirmasi bahwa jarak Manhattan dengan k=2 merupakan pilihan yang optimal dalam mendefinisikan struktur klaster, karena mayoritas sampel memiliki nilai SC yang positif dan cukup tinggi. Klasterisasi ini juga lebih stabil dibandingkan beberapa metrik lainnya yang sebelumnya memiliki nilai silhouette yang jauh lebih rendah. Namun, distribusi yang sangat tidak seimbang antara klaster 0 dan klaster 1 bisa menjadi indikasi bahwa data memiliki struktur yang secara alami condong ke satu kelompok besar dengan satu kelompok kecil yang lebih terisolasi.



Gambar 6 Grafik sebaran data pada klaster k = 2 dengan *manhattan distance*

Grafik yang ditampilkan pada Gambar 6 menunjukkan sebaran jumlah sampel dalam masing-masing klaster berdasarkan hasil klasterisasi K-Means dengan k=2 menggunakan metrik jarak Manhattan. Dari visualisasi ini, terlihat bahwa klaster 0 mendominasi dengan jumlah sampel yang jauh lebih besar dibandingkan klaster 1, yang hanya memiliki sedikit anggota. Ketimpangan ini mengindikasikan bahwa data yang digunakan memiliki struktur yang cenderung mengelompok ke dalam satu klaster utama, sementara klaster lainnya berisi sejumlah kecil sampel yang lebih terisolasi. Jika dikaitkan dengan silhouette plot sebelumnya, pola ini semakin memperkuat temuan bahwa klaster 1 memiliki nilai SC yang tinggi dan stabil, yang berarti objek dalam klaster ini memiliki jarak yang jelas terhadap klaster lain. Sebaliknya, klaster 0 berisi mayoritas sampel, tetapi memiliki variasi dalam nilai SC, dengan beberapa objek kemungkinan berada di batas antara klaster atau memiliki kedekatan dengan klaster lainnya. Namun, karena rata-rata SC masih tinggi (0,7191), bisa disimpulkan bahwa meskipun ada ketimpangan jumlah sampel antar klaster, pemisahan antar klaster cukup jelas dan valid. Hasil klasterisasi menunjukkan bahwa Klaster 0 didominasi peserta dengan metode suntik dan pil, sedangkan Klaster 1 lebih banyak peserta dengan metode IUD. Implikasi ini penting bagi pengambil kebijakan, karena klasterisasi dapat membantu menentukan strategi penyuluhan dan distribusi alat kontrasepsi yang lebih tepat sasaran.

4. KESIMPULAN

Penelitian ini menegaskan bahwa pemilihan ukuran jarak (*distance measure*) memiliki pengaruh signifikan terhadap kualitas hasil klasterisasi data peserta program Keluarga Berencana (KB) di

Desa Kalirejo. Berdasarkan hasil eksperimen menggunakan delapan jenis ukuran jarak, ditemukan bahwa algoritma K-Means dengan Manhattan distance pada k=2 memberikan hasil terbaik dengan Silhouette Coefficient (SC) sebesar 0,7191, menandakan struktur klaster yang sangat kuat dan terpisah dengan baik. Sebaliknya, ukuran jarak Cosine dan Canberra menunjukkan kinerja terendah dengan nilai SC masing-masing sebesar 0,2166 dan 0,2664, yang menandakan bahwa kedua ukuran tersebut kurang cocok digunakan pada dataset dengan karakteristik seperti ini. Temuan ini memperkuat bukti empiris bahwa pemilihan metrik jarak yang sesuai harus mempertimbangkan distribusi dan skala data. Hasil penelitian memberikan kontribusi metodologis dalam penerapan K-Means untuk data demografis, khususnya dalam konteks kebijakan dan perencanaan program KB berbasis data. Implikasi praktisnya adalah pemerintah daerah dan BKKBN dapat memanfaatkan pendekatan klasterisasi dengan ukuran jarak optimal untuk mengidentifikasi pola penggunaan kontrasepsi dan merancang strategi intervensi yang lebih efektif. Penelitian lanjutan disarankan untuk mengeksplorasi algoritma klasterisasi lain seperti K-Medoids atau DBSCAN guna membandingkan ketahanan hasil terhadap noise dan bentuk data yang kompleks.

DAFTAR PUSTAKA

Andriyani, W., Anshori, M., Normawati, D., Pradini, R. S., Zaenudin, M., Harisuddin, M. I., Haris, M. S., Astuty Sitinjak A. A., & Kusuma, W. T. (2024). *Matematika pada Kecerdasan Buatan*. Tohar Media.

Argiento, R., Filippi-Mazzola, E., & Paci, L. (2025). Model-Based Clustering of Categorical Data Based on the Hamming Distance. *Journal of the American Statistical Association*, *120*(550), 1178–1188. https://doi.org/10.1080/01621459.2024.2402568

Ghosh, A. (2022). Prediction of Diabetes Using Random Forest and XGBoost Classifiers. *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)*, *12*(1), 19–28.

Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, *9*(1), 606–620. https://doi.org/10.35957/jatisi.v9i1.1516

Hutagalung, J., & Sonata, F. (2021). Penerapan Metode K-Means untuk Menganalisis Minat Nasabah. *Jurnal Media Informatika Budidarma*, *5*(3), 1187–1194. https://doi.org/10.30865/mib.v5i3.3113

Idrus, A., Tarihoran, N., Supriatna, U., Tohir, A., Suwarni, S., & Rahim, R. (2022). Distance Analysis Measuring for Clustering Using K-Means and Davies Bouldin Index Algorithm. *TEM Journal*, *11*(4), 1871–1876. https://doi.org/10.18421/TEM114-55

Jollyta, D., Prihandoko, P., Priyanto, D., Hajjah, A., & Nora Marlim, Y. (2023). Comparison of Distance Measurements Based on k-Numbers and Its Influence to Clustering. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, *23*(1), 93–102. https://doi.org/10.30812/matrik.v23i1.3078

Kumala, A. A. S. P. A., & Rahning Putri, L. A. A. (2023). Measure Comparison Distance on K-Means Clustering for Grouping Music on Mood. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, *11*(4), 663–674. https://doi.org/10.24843/JLK.2023.v11.i04.p03

Lino, M., Jedo, A., & Adam, C. V. (2021). Identifikasi Faktor-Faktor yang Mempengaruhi Pengambilan Keputusan Pasangan Usia Subur dalam Mengikuti Program KB (Studi Kasus di Desa Leraboleng Kecamatan Titehena Kabupaten Flores Timur). *Jurnal Administrasi dan Demokrasi*, *1*(2), 101–123. https://doi.org/10.35508/jad.v1i2.5599

Maharani, S., & Yotenka, R. (2024). Pengelompokan Kecamatan di Daerah Istimewa Yogyakarta Berdasarkan Jumlah Pengguna Alat Kontrasepsi Tahun 2022 dengan K-Medoids Cluster. *Emerging Statistics and Data Science Journal*, *2*(2), 222–237. https://doi.org/10.20885/esds.vol2.iss.2.art16

Maulana, P. F. (2020). Upaya Dinas Kesehatan dan Keluarga Berencana dalam Pelaksanaan Kebijakan Keluarga Berencana di Kelurahan Bontang Lestari Kota Bontang. *EJournal Ilmu Pemerintahan*, *8*(3), 741–754. https://ejournal.ip.fisip-unmul.ac.id/site/wp-content/uploads/2021/01/Pungki

Munawar, F., Utami, A. S. D., & Manurung, S. B. T. (2024). Klasterisasi Daerah Peserta KB Aktif di Kabupaten Asahan Menggunakan Metode K-Means. *J-Com (Journal of Computer)*, *4*(1), 58–67. https://doi.org/10.33330/j-com.v4i1.3047

Muttaqin, J. A., Harlina, S., S, W., Hakim, L., Anshori, M., Ambarwari, A., Kaunang, F. J., Sandag, G. A., Harizahayu, M. G. F., Ruslau, M. F. V, Prasetio, A., Nasir, K. R., & SIregar, M. N. H. (2023). *Data Science dan Pembelajaran Mesin*. Yayasan Kita Menulis.

Muttaqin, W. W. W., Munsarif, M., Mandias, G. F., Pungus, S. R., Widarman, A., Hapsari, W. K., Hardiyanti, S. A., Fatkhudin, A., Pasnur, B. E. F., Anshori, M. S., & Saputra, N. (2023). *Pengenalan Data Mining*. Yayasan Kita Menulis.

Ningrum, A. V. F. F., Anshori, M., & Pradini, R. S. (2025). Klasterisasi Peserta KB Aktif di Desa Kalirejo Lawang Menggunakan Metode K-Means. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, *6*(1), 729–741. https://doi.org/10.35870/jimik.v6i1.1273

Pangestu, M. S., & Fitriani, M. A. (2022). Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means. *Sainteks*, *19*(2), 141–155. https://doi.org/10.30595/sainteks.v19i2.14495

Pratistha, R. N., & Kristianto, B. (2024). Implementasi Algoritma K-Means dalam Klasterisasi Kasus Stunting pada Balita di Desa Randudongkal. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, *5*(2), 1193–1205. https://doi.org/10.35870/jimik.v5i2.634

Rawal, K., Parthvi, A., Choubey, D. K., & Shukla, V. (2021). Prediction of Leukemia by Classification and Clustering Techniques. In P. Kumar, Y. Kumar, & M. A. Tawhid (Eds.), *Machine Learning, Big Data, and IoT for Medical Informatics* (pp. 275–295). Elsevier. https://doi.org/10.1016/B978-0-12-821777-1.00003-3

Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 747–748. https://doi.org/10.1109/DSAA49011.2020.00096

Shapcott, Z. (2024). *An Investigation into Distance Measures in Cluster Analysis*. http://arxiv.org/abs/2404.13664

Solikhun, S., Siregar, M. R., Pujiastuti, L., Wahyudi, M., & Kurniawan, D. (2025). Comparison of Manhattan and Chebyshev Distance Metrics in Quantum-Based K-Medoids Clustering. *SISTEMASI*, *14*(4), 1562–1572. https://doi.org/10.32520/stmsi.v14i4.5193

Sumarsih, S. (2023). Hubungan Karakteristik Ibu Nifas Terhadap Pemilihan Metode Kontrasepsi Pascasalin di Puskesmas Selopampang Kabupaten Temanggung. *Sinar: Jurnal Kebidanan*, *5*(1), 1–14. https://doi.org/10.30651/sinar.v5i1.17321

Telsiz Kayaoğlu, G. İ., & Eroğlu, M. (2024). Farklı Uzaklık Fonksiyonlarının Spektral Kümeleme Algoritmasının Performansına Etkisi. *Deu Muhendislik Fakultesi Fen ve Muhendislik*, *26*(77), 237–241. https://doi.org/10.21205/deufmd.2024267706

Tifannii, W. F., Mayasari, M., & Maulana, R. (2020). Implementasi Program Keluarga Berencana (KB) Dalam Upaya Menekan Pertumbuhan Penduduk di Kelurahan Sumur Batu Kecamatan Bantar Gebang Kota Bekasi. *Jurnal Imiah Ilmu Administrasi*, *7*(3), 525–540. https://doi.org/10.25157/dinamika.v7i3.4348

Wala, J., Herman, H., & Umar, R. (2024). Implementasi K-Means Clustering pada Pengelompokan Pasien Penyakit Jantung. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *9*(3), 205–216. https://doi.org/10.14421/jiska.2024.9.3.205-216

Wurdianarto, R. S., Novianto, S., & Rosyidah, U. (2014). Perbandingan Euclidean Distance dengan Canberra Distance pada Face Recognition. *Techno.COM*, *13*(1), 31–37. https://doi.org/10.33633/tc.v13i1.539