

Article Identity

Volume 1 Issue 1 Pages 19–26

October 2026

E-ISSN xxxx-xxxx

Article History

Received 30 July 2026

Revised 28 August 2026

Accepted 08 September 2026

Published 14 September 2026

Keywords

Sports Infrastructure; Village
Potential Statistics; Clustering;
Fuzzy C-Means; Silhouette
Coefficient

Copyright © 2026 Author(s).

Editorial office: Journal of Statistics and
Data Analytics, PT. Gammabeta Edu and
Tech.

To Cite this Article: Muhammad Arya
Syaputra et al. Analisis Pengelompokan
Sarana Olahraga Pedesaan Menggunakan
Algoritma Fuzzy C-Means. Journal of
Statistics and Data Analytics.
2026;1(1):19–26.

RESEARCH ARTICLE

Analisis Pengelompokan Sarana Olahraga Pedesaan Menggunakan Algoritma Fuzzy C-Means

Muhammad Arya Syaputra¹, Amelia Reza Safitri², Bella Andika Fatma
Fitriya³, Dian Candra Rini Novitasari^{4*}, Adelia Rova Chumairo⁵

^{1,2,3,4,5}Program Studi Matematika, Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya,
Indonesia

Corresponding author: diancrini@uinsa.ac.id

Abstract. Sports infrastructure is a crucial component that supports community physical activities, improves public health, and contributes to the development of national sports achievements. This study utilizes data from the 2024 Village Potential Statistics (PODES). The dataset includes seven types of sports facilities and one supporting variable. The facilities examined in this study consist of soccer fields, volleyball courts, badminton courts, basketball courts, tennis courts, futsal courts, swimming pools, and one additional supporting variable. The purpose of this study is to classify provinces in Indonesia based on the availability of village sports infrastructure using the Fuzzy C-Means method. The quality of the clustering results was evaluated using the Silhouette Coefficient, which produced an average value of 0.70043. This result indicates that the clusters formed exhibit a high degree of within-cluster homogeneity and clear separation between clusters. Therefore, the clustering results can be categorized as having high quality and meeting statistical acceptance criteria.

Abstrak. Infrastruktur olahraga merupakan komponen krusial yang mendukung aktivitas fisik masyarakat, peningkatan derajat kesehatan publik, dan pengembangan prestasi olahraga nasional. Riset ini memakai data dari Statistik Potensi Desa (PODES) Tahun 2024. Di dalamnya ada tujuh jenis fasilitas olahraga dan satu variabel pendukung yang diamati. Infrastruktur olahraga yang digunakan dalam penelitian ini meliputi lapangan sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, futsal, kolam renang, serta satu variabel pendukung lainnya. Tujuan penelitian ini adalah untuk mengelompokkan provinsi Indonesia berdasarkan tingkat ketersediaan infrastruktur olahraga desa dengan menggunakan logika Fuzzy C-Means. Evaluasi kualitas pengelompokan dilakukan melalui pengukuran silhouette coefficient yang menghasilkan nilai rata-rata sebesar 0.70043. Hasil tersebut menunjukkan bahwa cluster yang terbentuk memiliki tingkat homogenitas dalam kelompok dan perbedaan antar kelompok yang baik. Dengan demikian, hasil pengelompokan dapat dikategorikan memiliki kualitas yang tinggi serta memenuhi kriteria penerimaan secara statistik.

1. PENDAHULUAN

Olahraga adalah salah satu aktivitas penting dalam menunjang kehidupan agar tubuh tetap sehat dan bugar [1]. Olahraga memiliki peran penting dalam meningkatkan kesehatan, kebugaran jasmani, kualitas hidup, serta produktivitas masyarakat [2]. Ketersediaan sarana dan prasarana olahraga memadai di setiap wilayah serta menjadi faktor pendukung dalam pelaksanaan aktivitas olahraga secara berkelanjutan dan pembinaan prestasi olahraga [3]. Infrastruktur olahraga yang memadai juga berkontribusi terhadap peningkatan partisipasi masyarakat dalam berolahraga dan mendukung pembangunan sumber daya manusia yang berkualitas.

Menurut hasil publikasi Potensi Desa (PODES) tahun 2024 yang dikeluarkan oleh Badan Pusat Statistik (BPS), ketersediaan infrastruktur olahraga di tingkat desa menunjukkan perbedaan yang cukup signifikan antarwilayah [28]. Data tersebut mencatat bahwa jumlah fasilitas olahraga seperti lapangan bola voli, sepak bola, basket, tenis, futsal, bulu tangkis, dan kolam renang tidak tersebar merata di seluruh wilayah Indonesia. Sebagian besar fasilitas olahraga masih terkonsentrasi di wilayah tertentu, sementara daerah lainnya masih mengalami kekurangan infrastruktur dasar tersebut [5]. Sebagian desa memiliki fasilitas olahraga yang relatif lengkap, sedangkan desa lainnya masih memiliki keterbatasan sarana olahraga yang memadai. Kondisi tersebut mengindikasikan adanya kesenjangan ketersediaan infrastruktur olahraga antarwilayah, sehingga diperlukan identifikasi karakteristik desa guna mendukung perencanaan pembangunan yang lebih tepat sasaran.

Ketersediaan fasilitas olahraga berpengaruh terhadap aktivitas fisik masyarakat, berdampak pada tingkat kebugaran jasmani, pembinaan prestasi, dan kualitas hidup [6]. Aksesibilitas fasilitas olahraga memiliki peran penting terhadap tingkat partisipasi masyarakat dalam kegiatan olahraga rekreasi. Ketersediaan fasilitas olahraga lebih banyak ditemukan di kota-kota besar, sedangkan wilayah pedesaan dan kawasan 3T (tertinggal, terluar, dan terdepan) masih menghadapi keterbatasan infrastruktur olahraga, yang berdampak langsung terhadap rendahnya partisipasi olahraga di daerah [7]. Kondisi ini turut berdampak pada kebugaran jasmani masyarakat dan indeks kebugaran jasmani nasional tercatat terus mengalami penurunan selama hampir dua dekade terakhir. Diperlukan suatu metode pengelompokan desa atau kelurahan

berdasarkan tingkat ketersediaan fasilitas olahraga untuk penentuan prioritas pemerataan pembangunan sarana olahraga lebih efektif dan tepat sasaran. Proses pengelompokan dapat dilakukan dengan berbagai metode diantaranya, *K-Means* dan *Fuzzy C-Means*.

Penelitian terdahulu menggunakan metode *clustering* untuk mengelompokkan pola dan kemiripan karakteristik antarwilayah sebagai dasar perencanaan pembangunan [8], [9]. Salah satu studi membandingkan metode *K-Means* dan *Self-Organizing Maps (SOM)* untuk pengelompokan provinsi di Indonesia berdasarkan data Potensi Desa [9]. Studi lainnya menerapkan analisis *cluster* pada pengelompokan desa di Kecamatan Pariaman Timur berdasarkan Indeks Desa Membangun 2023 [8]. Namun fokusnya pada provinsi secara umum dan belum secara umum mengkaji fasilitas olahraga. Dalam penelitian ini digunakan metode *clustering* berbasis logika *fuzzy* sebagai dasar metode *Fuzzy C-Means*. *Clustering* merupakan teknik analisis data yang bertujuan mengelompokkan objek berdasarkan tingkat kemiripan karakteristik yang dimiliki. Melalui metode tersebut, setiap objek studi dapat memiliki nilai keanggotaan pada lebih dari satu kelompok dengan nilai antara 0 hingga 1, sehingga mampu merepresentasikan karakteristik data yang heterogen secara lebih fleksibel, dengan perolehan nilai *silhouette coefficient* sebesar 0.7 yang menunjukkan kualitas pengelompokan berada pada kategori baik sehingga sesuai digunakan untuk mengelompokkan wilayah berdasarkan ketersediaan infrastruktur olahraga desa.

Penelitian sebelumnya menunjukkan bahwa algoritma *Fuzzy C-Means* efektif digunakan untuk mengelompokkan data yang memiliki karakteristik heterogen. Studi pada pengelompokan data keluarga menggunakan *Fuzzy C-Means* menunjukkan bahwa metode ini mampu menghasilkan *cluster* yang representatif berdasarkan kemiripan atribut [10]. Studi lain menganalisis pengaruh pemilihan nilai *fuzziness exponent* terhadap hasil pengelompokan provinsi di Indonesia berdasarkan indikator pembangunan ekonomi dan menunjukkan bahwa pemilihan parameter yang tepat dapat meningkatkan kualitas hasil pengelompokan [11]. Berdasarkan keberhasilan penerapan metode *Fuzzy C-Means* pada berbagai penelitian tersebut, metode ini dinilai sesuai untuk digunakan dalam pengelompokan wilayah berdasarkan ketersediaan infrastruktur olahraga desa. Nilai *Silhouette Coefficient* yang di peroleh pada penelitian ini adalah 0.7 sehingga, dapat diinterpretasikan bahwa hasil clustering berada dalam kategori baik dan mampu menggambarkan kemiripan karakteristik antarwilayah dengan cukup jelas.

Dengan mempertimbangkan hasil penelitian terdahulu, penelitian ini bertujuan untuk mengelompokkan wilayah berdasarkan ketersediaan infrastruktur olahraga desa yang meliputi lapangan sepak bola, lapangan bola voli, lapangan bulu tangkis, lapangan basket, lapangan tenis, lapangan futsal, dan kolam renang. Penelitian ini menggunakan evaluasi hasil *clustering* menggunakan *silhouette coefficient*. Hasil penelitian diharapkan mampu memberikan gambaran karakteristik infrastruktur olahraga desa sebagai dasar dalam penyusunan kebijakan pembangunan sarana olahraga yang lebih efektif merata dan tepat sasaran.

2. KAJIAN TEORI

2.1. Olahraga

Olahraga menjadi bagian penting untuk menjaga kebugaran tubuh dan menjaga kestabilan mental [12]. Kegiatan jasmani yang dikerjakan secara rutin memiliki peran penting dalam pencegahan berbagai penyakit menular, seperti diabetes melitus, hipertensi, gangguan kesehatan mental, dan penyakit kardiovaskular. Selain itu, rajin berolahraga juga dapat meningkatkan kualitas hidup sehingga olahraga memiliki peran penting bagi kesehatan. Kondisi fisik yang bugar adalah komponen penting dalam mendukung kinerja olahraga. Kebugaran yang prima menjadi fondasi bagi peningkatan kemampuan teknik, strategi serta ketahanan tubuh seseorang dalam menjalankan aktivitas olahraga dengan optimal [13].

2.2. Analisis Korelasi

Korelasi adalah ukuran statistik yang menunjukkan sejauh mana dua variabel atau lebih memiliki hubungan ketertarikan satu sama lain [14]. Normalisasi data itu untuk mengubah rentang nilai data yang ada agar nilai data tersebut berada dalam rentang nilai yang sama. Analisis korelasi ini sangat penting digunakan untuk meminimalisir penumpukan data yang dapat menyebabkan data menjadi tidak teratur [15]. Hal ini terjadi karena variabel-variabel yang sangat berkaitan satu dengan yang lain dan dapat memengaruhi pembentukan *cluster* secara tidak seimbang. Jika nilai korelasi mendekati 1, maka nilai tersebut menunjukkan hubungan yang kuat, sedangkan korelasi negatif menandakan adanya hubungan yang berlawanan arah antar variabel [16].

2.3. Normalisasi

Normalisasi data merupakan teknik prapemrosesan (*preprocessing*) yang digunakan untuk menyeragamkan skala nilai antar variabel [17]. Perbedaan rentang nilai antar variabel berpotensi memengaruhi hasil analisis. Proses normalisasi mentransformasikan data ke dalam rentang tertentu tanpa mengubah karakteristik dan pola hubungan pada data asli. Tahapan ini menjadi bagian penting dalam analisis *clustering*. Metode *min max scaler* merupakan proses data yang menggunakan nilai maksimum dan minimum dari suatu atribut untuk mengonversi data ke dalam rentang [0, 1] sehingga menghasilkan perbandingan antar data yang seimbang. Baik sesudah atau sebelum proses normalisasi. Persamaan (1) di bawah ini merupakan formulasi *Min-Max Scaler* [18]

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Dengan x' merupakan data yang telah dinormalisasi, x merupakan data asli, $x_{\max}$ merupakan nilai maksimum pada atribut, $x_{\min}$ merupakan nilai minimum pada atribut.

2.4. Fuzzy C-Means

Fuzzy C-Means (FCM) adalah algoritma pengelompokan data yang pertama kali dipublikasikan oleh Jim Bezdek pada tahun 1981 [19]. Dalam metode ini, nilai keanggotaan dimanfaatkan untuk menentukan tingkat keterkaitan suatu data pada *cluster* tertentu, sehingga satu data dapat menjadi anggota lebih dari satu *cluster*. Output yang dihasilkan mencakup sentroid *cluster* dan nilai keanggotaan setiap data. Pada metode *Fuzzy C-Means*, sentroid *cluster* ditentukan melalui proses iterasi berulang hingga diperoleh posisi yang optimal. Penentuan keanggotaan data pada suatu *cluster* didasarkan pada nilai keanggotaan yang dimiliki. Metode ini memiliki kelebihan dalam membentuk pusat *cluster* yang mampu mengakomodasi data dengan karakteristik yang saling beririsan antar kelompok [20]. Tahapan algoritma dari *Fuzzy C-Means*, dijelaskan pada bagian berikut [21].

1. Input seluruh data yang akan di *cluster* x , dalam bentuk matriks berukuran $n \times m$ (n = total jumlah data, m = jumlah atribut data).

- x_{ij} = nilai data pada sampel ke- i ($i = 1, 2, 3, \dots, n$), dan atribut ke- j ($j = 1, 2, \dots, m$).
2. Tentukan parameter yang dibutuhkan, yaitu:
 - a. Jumlah *cluster* (C)
 - b. Bobot pangkat (w)
 - c. Batas maksimum iterasi ($MaxIter$)
 - d. Ambang error terkecil (ε)
 - e. Nilai fungsi objektif awal ($P_0 = 0$)
 - f. Indeks iterasi awal ($t = 1$)
 3. Bangkitkan bilangan acak μ_{ik} ($i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$) yang merupakan elemen-elemen matriks partisi awal U . Kemudian, hitung jumlah setiap kolom dengan formulasi (2)

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (2)$$

Untuk indeks:

μ_{ik} = Derajat keanggotaan terhadap *cluster*

Q_i = Jumlah nilai derajat keanggotaan per kolom = 1 dengan $i = 1, 2, \dots, n$.

Selanjutnya tentukan nilai matriks partisi awal menggunakan formulasi (3).

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (3)$$

4. Hitung titik tengah *cluster* ke- k : V_{kj} , dengan $k = 1, 2, \dots, c$ dan $j = 1, 2, \dots, n$ memakai formulasi (4).

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w X_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (4)$$

Untuk indeks:

V_{kj} = pusat *cluster*

x_i = parameter ke- i

5. Kalkulasi nilai objektif pada iterasi ke- t (P_t) dengan rumus (5).

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \quad (5)$$

Dimana P_t adalah nilai fungsi objektif iterasi ke- t .

6. Perubahan matriks nilai keanggotaan menggunakan formulasi (6).

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-1/(w-1)}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-1/(w-1)}} \quad (6)$$

Untuk indeks $i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$

7. Periksa syarat pemberhentian dengan kriteria berikut :
 - a. Jika $(P_t - (P_t - 1) < \varepsilon)$ atau $(t > MaxIter)$, maka proses dihentikan
 - b. Sebaliknya, perbarui nilai : $t = t + 1$, lalu kembali ke langkah ke-4

2.5. Denormalisasi

Denormalisasi database merupakan teknik dalam perancangan basis data yang digunakan dengan cara menambahkan redundansi data ke dalam struktur tabel yang sebelumnya telah dinormalisasi [22]. Tujuan utama dari denormalisasi adalah untuk meningkatkan performa sistem [23]. Data penelitian yang telah dinormalisasi akan dikonversi kembali ke nilai aslinya melalui proses denormalisasi agar serupa dengan data riil. Teknik ini memudahkan perbandingan antara hasil perhitungan sistem dengan data riil menggunakan formulasi (7) [24]:

$$x_i = x' (x_{\max} - x_{\min}) + x_{\min} \quad (7)$$

Keterangan :

x_i = Data aktual yang dinormalisasikan

x' = Data output

$x_{\min}$ = Nilai minimum yang diaktualisasi

$x_{\max}$ = Nilai maksimum yang diaktualisasi

2.6. Silhouette Coefficient

Silhouette diterapkan untuk dapat mengevaluasi mutu serta kepadatan dari tiap-tiap kelompok data dengan cara menghitung skor bagi tiap objek data berdasarkan jarak rata-rata ke titik lain dalam *cluster* yang sama dan jarak rata-rata ke titik di *cluster* tetangga terdekat [25]. Nilai yang dihasilkan berkisar antara -1 hingga 1, dengan nilai yang mendekati 1 menunjukkan *cluster* yang terpisah dengan baik, nilai yang mendekati 0

menunjukkan *cluster* yang tumpang tindih, dan nilai negatif menunjukkan potensi kesalahan klasifikasi [26]. Indeks *silhouette* keseluruhan diperoleh dengan merata-ratakan skor seluruh titik data, sehingga memberikan penilaian keseluruhan terhadap struktur pengelompokan. *silhouette* dapat dilakukan dengan rumus (8).

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (8)$$

Keterangan:

S_i = *Silhouette Coefficient*.

a_i = Rata-rata jarak dari objek i terhadap seluruh objek dalam kelompok yang sama.

b_i = Nilai rata-rata jarak terkecil antara objek i dengan objek di kelompok lain.

Tabel 1. Penjelasan Subjektif Silhouette Coefficient

Nilai Silhouette Coefficient	Definisi
0.71–1.00	Hasil <i>clustering</i> sangat baik
0.51–0.70	Hasil <i>clustering</i> cukup baik
0.26–0.50	Hasil <i>clustering</i> kurang optimal
≤ 0.25	Hasil <i>clustering</i> buruk

Berdasarkan Tabel 1 menunjukkan bahwa semakin mendekati 1 nilai *silhouette coefficient*, maka kualitas pengelompokan dianggap semakin optimal, sebaliknya nilai yang mendekati 0 mengindikasikan adanya tumpang tindih antar data, yang menandakan hasil *klasterisasi* kurang efektif [27].

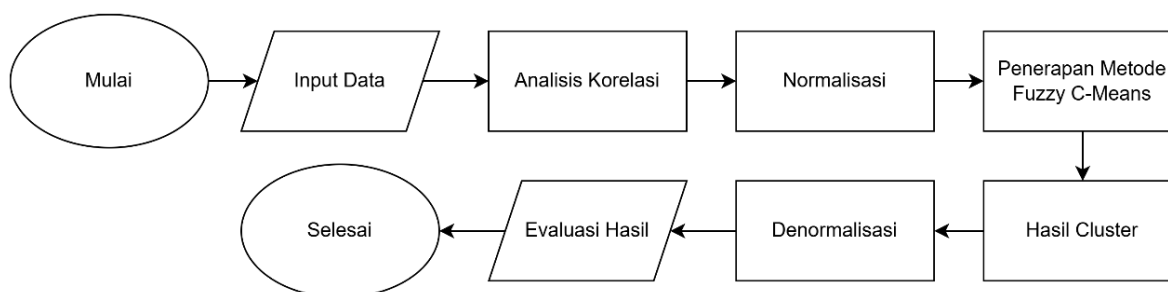
3. METODOLOGI

Penelitian ini memanfaatkan data sekunder yang telah diolah dan dipublikasikan untuk seluruh provinsi di Indonesia oleh Badan Pusat Statistik (BPS) [28]. Salah satu contoh data yang digunakan adalah data desa/kelurahan menurut ketersediaan lapangan olahraga pada tahun 2024 untuk seluruh provinsi di Indonesia.

Tabel 2. Sampel Data Penelitian

Provinsi	X1	X2	X3	X4	X5	X6	X7	X8
ACEH	2760	3931	1088	125	87	597	39	6516
SUMATERA UTARA	2065	3233	2183	127	134	812	211	6113
SUMATERA BARAT	893	1021	770	174	62	427	68	1265
RIAU	1491	1785	1326	147	69	636	108	1867
JAMBI	1164	1395	1244	114	38	412	58	1585
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PAPUA BARAT DAYA	236	386	32	29	7	80	11	1018
PAPUA	380	670	93	46	21	85	9	1029
PAPUA SELATAN	291	353	48	20	4	33	2	690
PAPUA TENGAH	321	576	52	19	12	33	11	1208
PAPUA PEGUNUNGAN	415	939	25	12	7	15	8	2634

Dalam Tabel 2 terdapat nama-nama provinsi di Indonesia. Variabel X1 menjelaskan jumlah lapangan sepak bola tahun 2024, X2 menjelaskan jumlah lapangan bola voli tahun 2024, X3 menjelaskan jumlah lapangan bulu tangkis tahun 2024, X4 menjelaskan jumlah lapangan bola basket tahun 2024, X5 menjelaskan jumlah lapangan tenis lapangan tahun 2024, X6 menjelaskan jumlah lapangan futsal tahun 2024, X7 menjelaskan jumlah kolam renang tahun 2024, dan X8 menjelaskan jumlah desa pada setiap provinsi.



Gambar 1. Flowchart Penelitian

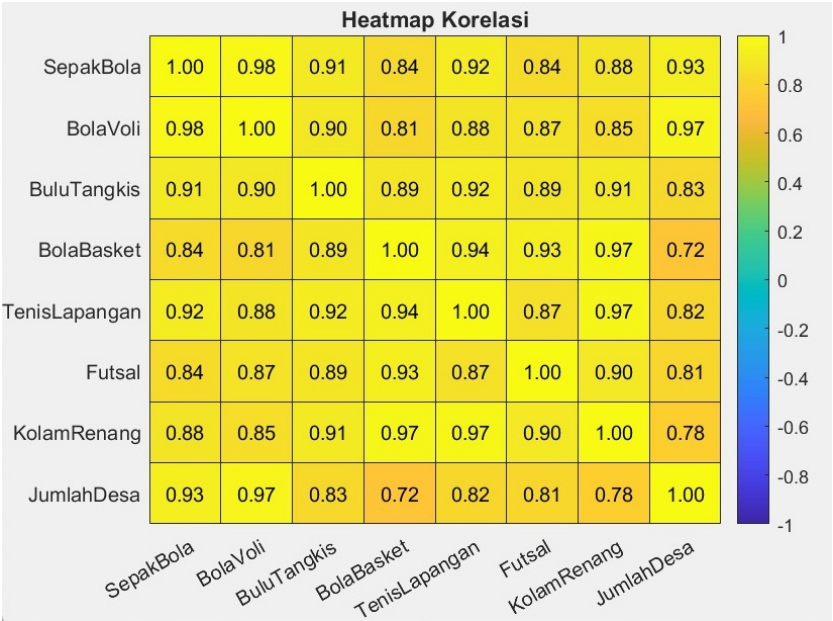
Alur penelitian di atas menggambarkan rangkaian proses penelitian yang dilaksanakan secara sistematis. Tahap awal penelitian dilakukan melalui pengumpulan data ketersediaan infrastruktur olahraga yang diperoleh dari Badan Pusat Statistik terkumpul. Langkah awal analisis dimulai dengan melihat tingkat kedekatan tiap variabel melalui uji korelasi. Untuk menghindari ketimpangan nilai saat pengujian, tahapan normalisasi data juga diterapkan untuk menyelaraskan seluruh skala variabel yang diperoleh sehingga tidak terjadi dominasi nilai pada proses analisis. Variabel yang memiliki hubungan kuat kemudian digabungkan untuk mengurangi redundansi informasi dalam data. Tahap berikutnya adalah penerapan metode Fuzzy C-Means untuk membentuk kelompok-kelompok berdasarkan karakteristik infrastruktur olahraga yang dimiliki. Hasil clustering yang diperoleh selanjutnya melalui proses denormalisasi agar nilai centroid kembali ke skala aslinya dan lebih mudah diinterpretasikan. Pada tahap akhir, kualitas cluster dievaluasi menggunakan silhouette coefficient untuk mengukur tingkat ketepatan

pengelompokan yang terbentuk. Berdasarkan hasil evaluasi, diperoleh nilai sebesar 0.70043, yang berarti setiap objek dalam cluster sudah sangat serupa dan terpisah secara tegas dari cluster yang berbeda. Alur penelitian ini disajikan untuk memudahkan pembaca terhadap alur dan tahapan penelitian yang dijalankan.

4. HASIL DAN PEMBAHASAN

4.1. Analisis Korelasi

Tahap awal analisis melibatkan pengujian hubungan antar variabel menggunakan visualisasi heatmap. Matriks korelasi dihitung untuk seluruh 8 variabel yang terdiri dari berbagai sarana olahraga (Bola Basket, Bola Voli, Sepak Bola, Futsal, Bulu Tangkis, Tenis Lapangan, Kolam Renang) serta Jumlah Desa di 38 provinsi Indonesia. Hasil analisis ini menggambarkan sejauh mana keterkaitan linier di antara indikator-indikator tersebut.



Gambar 2. Matriks Korelasi Antar Variabel

Sumber: Hasil Analisis, 2024

Dari hasil analisis korelasi, tampak bahwa seluruh variabel memiliki hubungan yang sangat kuat dan saling terkait satu sama lain. Korelasi tertinggi terlihat antara Sepak Bola dan Bola Voli dengan nilai 0.98, serta antara Bola Basket dengan Kolam Renang dan Tenis Lapangan yang masing-masing bernilai 0.97. Hal ini menunjukkan adanya kesamaan pola distribusi fasilitas olahraga antarprovinsi yang sangat tinggi. Korelasi terendah ditemukan antara Jumlah Desa dan Bola Basket dengan nilai 0.72, serta antara Jumlah Desa dan Kolam Renang sebesar 0.78. Korelasi tersebut tergolong rendah dibandingkan pasangan variabel lainnya, nilai-nilai tersebut masih menunjukkan korelasi positif yang cukup kuat. Seluruh pasangan variabel memiliki korelasi di atas 0.70, mengindikasikan tidak ada variabel yang bersifat independen satu sama lain. Kondisi ini menandakan adanya kemungkinan multikolinearitas. Hal ini perlu diperhatikan saat melanjutkan analisis dengan teknik pengelompokan *Fuzzy C-Means*.

4.2. Algoritma Fuzzy C-Means

Tahap selanjutnya melibatkan pengelompokan data menggunakan teknik *Klasterisasi Fuzzy C-Means*. Sebelum menerapkan metode *Fuzzy C-Means (FCM)* tersebut, kita perlu melakukan proses normalisasi terlebih dahulu menggunakan pendekatan *Min-Max Scaler* guna mentransformasi data ke dalam skala [0, 1]. Dengan melakukan proses ini, kita dapat meningkatkan kinerja *klasterisasi* serta dapat meminimalkan fitur dengan gangguan yang tinggi dan relevansi yang rendah. Dilanjutkan dengan menentukan parameter awal. Disajikan dalam Tabel 3 di bawah ini:

Tabel 3. Pusat Cluster Terakhir

Cluster	Sepak Bola	Bola Voli	Bulu Tangkis	Bola Basket	Tenis Lapangan	Futsal	Kolam Renang	Jumlah Desa
1	0.2438	0.3241	0.2373	0.1460	0.1613	0.2693	0.1105	0.3556
2	0.0655	0.0834	0.0602	0.0588	0.0611	0.0715	0.0362	0.1012
3	0.8336	0.9063	0.7959	0.7535	0.9299	0.7373	0.8754	0.8991

Berdasarkan Tabel 3, cluster 3 memiliki nilai centroid paling besar untuk seluruh variabel, yaitu lapangan sepak bola (0.8336), bola voli (0.9063), bulu tangkis (0.7959), bola basket (0.7535), tenis lapangan (0.9299), futsal (0.7373), kolam renang (0.8754), dan jumlah desa (0.8991), sehingga diberi label Ketersediaan Fasilitas Olahraga Tinggi. Cluster 1 memiliki nilai centroid pada kategori sedang untuk lapangan sepak bola (0.2438), bola voli (0.3241), bulu tangkis (0.2373), bola basket (0.1460), tenis lapangan (0.1613), futsal (0.2693), kolam renang (0.1105), dan jumlah desa (0.3556), sehingga diberi label Ketersediaan Fasilitas Olahraga Sedang. Cluster 2 memiliki nilai centroid terendah pada lapangan sepak bola (0.0655), bola voli (0.0834), bulu tangkis (0.0602), bola basket (0.0588), tenis lapangan (0.0611), futsal (0.0715), kolam renang (0.0362), dan jumlah desa (0.1012), sehingga diberi label Ketersediaan Fasilitas Olahraga Rendah. Tabel 4 berikut merinci pembagian provinsi ke dalam beberapa cluster.

Tabel 4 menunjukkan hasil klasterisasi ketersediaan fasilitas olahraga per provinsi yang dibagi menjadi tiga tingkatan: Tinggi, Sedang, serta Rendah. Rincian sebaran hasil pengelompokan untuk tiap-tiap provinsi dapat dicermati pada Tabel 5.

Tabel 4. Cluster Ketersediaan Fasilitas Olahraga di Indonesia

No.	Provinsi	Ketersediaan Tinggi	Ketersediaan Sedang	Ketersediaan Rendah
1	Sumatera Utara	–	1	–
2	Aceh	–	1	–
3	Sumatera Barat	–	–	2
4	Jawa Tengah	3	–	–
5	Papua Pegunungan	–	–	2
⋮	⋮	⋮	⋮	⋮
34	Papua Selatan	–	–	2
35	Jawa Barat	3	–	–
36	Papua Barat Daya	–	–	2
37	Jawa Timur	3	–	–
38	Jambi	–	–	2

Tabel 5. Hasil Klasterisasi Ketersediaan Fasilitas Olahraga Seluruh Provinsi di Indonesia

Kategori	Persentase	Nama Provinsi
Ketersediaan Fasilitas Olahraga Tinggi	7.89%	Jawa Timur, Jawa Barat, dan Jawa Tengah
Ketersediaan Fasilitas Olahraga Sedang	23.68%	Sumatra Utara, Sumatra Selatan, Kalimantan Barat, Aceh, Riau, Banten, Sulawesi Selatan, Lampung, Nusa Tenggara Timur
Ketersediaan Fasilitas Olahraga Rendah	68.42%	Jambi, Maluku, Bengkulu, Sumatra Barat, Kep. Bangka Belitung, Papua Pegunungan, Papua Tengah, Kep. Riau, Gorontalo, DKI Jakarta, Bali, Nusa Tenggara Barat, Kalimantan Tengah, DI Yogyakarta, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Barat, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan

Hasil analisis Tabel 5 memperlihatkan bahwa infrastruktur olahraga di Indonesia belum merata hanya 7.89% dari seluruh provinsi, yaitu Jawa Tengah, Jawa Timur, dan Jawa Barat yang tergolong dalam kategori fasilitas tinggi. Kelompok Ketersediaan Fasilitas Olahraga Sedang mencakup 9 provinsi atau 23.68%, yaitu Sumatra Utara, Sumatra Selatan, Kalimantan Barat, Aceh, Lampung, Riau, Banten, Sulawesi Selatan, dan Nusa Tenggara Timur. Kelompok Ketersediaan Fasilitas Olahraga Rendah mencakup 26 provinsi atau 68.42%, yaitu Sumatera Barat, Jambi, Bengkulu, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, Nusa Tenggara Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua Selatan, Papua Tengah, Papua Pegunungan, dan Papua. Kondisi yang mendominasi kategori rendah ini memberikan gambaran bahwa secara kolektif, penyediaan sarana olahraga di berbagai provinsi masih tertinggal jauh jika dibandingkan dengan provinsi-provinsi di kategori sedang dan tinggi.

4.3. Evaluasi Hasil Klasterisasi (*Silhouette Coefficient*)

Tahap selanjutnya dilakukan uji silhouette coefficient untuk menilai kualitas hasil pengelompokan data. Nilai silhouette coefficient berkisar antara -1 dan 1. Semakin mendekati 1, semakin baik kualitas pengelompokan yang diperoleh, dan sebaliknya.

Berdasarkan perhitungan yang telah dilakukan dengan 3 kelompok (cluster), diperoleh nilai silhouette coefficient sebesar 0.70043. Nilai tersebut menunjukkan kualitas pengelompokan yang baik dan mendekati kategori sangat baik. Hasil ini juga menunjukkan bahwa anggota dalam setiap cluster memiliki tingkat kemiripan yang kuat dan terpisah dengan baik dari cluster lainnya. Dengan demikian, metode Fuzzy C-Means mampu mengelompokkan data sarana olahraga dan persebaran desa secara efektif.

4.4. Visualisasi Hasil Klaster

Hasil klasterisasi mengklasifikasikan tingkat ketersediaan fasilitas olahraga seluruh provinsi di Indonesia ke dalam tiga kategori utama: Ketersediaan Fasilitas Olahraga Tinggi (warna merah muda/pink), Ketersediaan Fasilitas Olahraga Sedang (warna ungu), dan Ketersediaan Fasilitas Olahraga Rendah (warna hijau).

Pemetaan ini memperlihatkan ketimpangan signifikan pemerataan infrastruktur olahraga antarwilayah Indonesia. Mayoritas wilayah luar Pulau Jawa didominasi warna hijau, mencerminkan ketersediaan fasilitas olahraga kategori rendah. Wilayah kategori rendah mencakup area sangat luas: sebagian Sumatra (Jambi), sebagian besar Kalimantan (Kalimantan Tengah dan Kalimantan Timur), seluruh wilayah Sulawesi, Kepulauan Maluku, Nusa Tenggara, DI Yogyakarta, dan Papua. Temuan ini mengindikasikan pembangunan, pengadaan fasilitas olahraga masih tersentralisasi, belum menjangkau wilayah luar secara merata.

Wilayah tingkat ketersediaan fasilitas olahraga sedang tersebar sporadis di beberapa provinsi strategis. Wilayah Pulau Sumatra mencakup Aceh, Sumatra Utara, Riau, Sumatra Selatan. Kategori sedang ditemukan juga di Kalimantan Barat, sebagian wilayah Nusa Tenggara, Maluku. Kondisi ini menunjukkan upaya pembangunan infrastruktur olahraga cukup berkembang di pusat pertumbuhan regional tersebut, meskipun belum mencapai tahap optimal.

Tingkat ketersediaan fasilitas olahraga tinggi (warna merah muda/pink) terkonsentrasi eksklusif di Pulau Jawa (Jawa Timur, Jawa Tengah, dan Jawa Barat). Fenomena tersebut selaras dengan kepadatan penduduk, kemajuan ekonomi regional, kelengkapan sarana publik wilayah tersebut.

Ketidakeimbangan visual ini menjadi tantangan besar pemerintah mewujudkan pola hidup sehat merata seluruh Indonesia. Diperlukan

Peta Indonesia



Gambar 3. Visualisasi Fuzzy C-Means

intervensi kebijakan fokus, terarah, guna mempercepat pembangunan, renovasi infrastruktur olahraga, khususnya wilayah klasifikasi rendah. Langkah ini penting mendukung potensi atlet daerah, kebugaran masyarakat umum dan untuk penelitian selanjutnya disarankan menggunakan variabel luas wilayah sebagai pengganti variabel jumlah desa dalam analisis ketersediaan fasilitas olahraga. Hal ini karena luas wilayah dinilai lebih mampu mempresentasikan cakupan area yang harus dilayani sehingga hasil klusterisasi dapat memberikan gambaran yang lebih akurat mengenai pemerataan fasilitas olahraga antarprovinsi di Indonesia.

5. KESIMPULAN

Penelitian ini berhasil mengidentifikasi pola ketersediaan infrastruktur olahraga desa di Indonesia melalui pendekatan Fuzzy C-Means yang menghasilkan tiga cluster wilayah (tinggi, sedang, dan rendah). Cluster tinggi didominasi oleh provinsi di Pulau Jawa, sedangkan kategori rendah mendominasi dengan proporsi sebesar 68.42% dari seluruh provinsi di Indonesia. Perolehan skor silhouette coefficient mencapai 0.70043. Angka ini membuktikan bahwa kualitas cluster yang diperoleh memiliki tingkat pemisahan dan homogenitas cukup baik. Penelitian ini menyajikan gambaran terkait ketersediaan infrastruktur olahraga antarprovinsi yang dapat dijadikan sebagai rujukan dalam penyusunan kebijakan pemerataan pembangunan fasilitas olahraga yang lebih terukur, merata, dan sesuai kebutuhan daerah. Penelitian ini juga menegaskan bahwa penggunaan metode Fuzzy C-Means mampu menjadi pendekatan yang efektif dalam mengklasifikasikan wilayah berdasarkan profil infrastruktur olahraga yang dimiliki.

DAFTAR PUSTAKA

- [1] I. Zubaida et al., “Olahraga Kesehatan: Memasyarakatkan Olahraga Untuk Peningkatan Kesehatan Sport Health: Promoting Exercise For Health Improvement Ilmu Keolahragaan, Fakultas Kedokteran, Universitas Sultan Ageng Tirtayasa,” vol. 1, no. 1, pp. 15–21, 2022.
- [2] E. Supriatna, D. Suryadi, and M. Haetami, “Analysis Endurance Profile (Vo2max) Of Women’s Volleyball Athletes: Yo-Yo Intermittent Test Level 1,” vol. 3, no. 1, pp. 12–19, 2023.
- [3] S. I. Keolahragaan, F. I. Olahraga, and U. N. Surabaya, “Pengaruh Olahraga Dan Model Latihan Fisik Terhadap Kebugaran Jasmani Remaja: Literature Review Dedy Pranata,” vol. 10, no. 2, pp. 107–116.
- [4] N. Utomo and M. H. Afiludin, “Ilmu Keolahragaan, Fakultas Ilmu Keolahragaan Dan Kesehatan, Universitas Negeri Yogyakarta Pendidikan Jasmani Sekolah Dasar, Fakultas Ilmu Keolahragaan Dan Kesehatan, Universitas Negeri Yogyakarta,” vol. 29, no. 2, pp. 36–45, 2023.
- [5] P. Sekolah, D. I. Kawasan, and P. Dan, “Jurnal Olahraga Dan Kesehatan Indonesia (JOKI): Analisis Perbandingan Sarana Dan Prasarana PJOK,” vol. 6, no. 1, pp. 80–87, 2025.
- [6] J. S. Tahun, T. N. Citrawati, and J. Ilyas, “Aplikasi Sistem Informasi Kebugaran,” vol. 8, no. April, pp. 1295–1303, 2024.
- [7] A. G. Permadi and M. R. Lubis, “Profil Sarana Dan Prasarana Pendidikan Jasmani, Olahraga, Dan Kesehatan Di Sekolah Menengah Pertama Negeri Se-Kota Mataram,” vol. 9, no. 1, pp. 716–724, 2023, doi: 10.58258/Jime.V9i1.4766/Http.
- [8] Y. Izamahendra and M. Subhan, “Desa Pariaman Timur Berdasarkan Indeks Desa Membangun 2023,” vol. 13, no. 2, pp. 204–211, 2025.
- [9] L. R. Iyohu, I. Djakaria, and L. O. Nashar, “Perbandingan Metode K-Means Clustering Dengan Self-Organizing Maps (SOM) Untuk Pengelompokan Provinsi Di Indonesia Berdasarkan Data Potensi Desa,” vol. 7, no. 2, pp. 195–206, 2023.
- [10] W. Cahyadi, E. Haerani, A. Nazir, and I. Iskandar, “TIN: Terapan Informatika Nusantara Analisis Algoritma Fuzzy C-Means Untuk Pengelompokan Data Keluarga,” vol. 6, no. 7, pp. 1181–1189, 2025, doi: 10.47065/Tin.V6i7.8981.
- [11] U. Sa, “Pythagoras: Jurnal Matematika Dan Pendidikan Matematika Seleksi Nilai Fuzziness Exponent Optimal Pada Algoritma Fuzzy C-Means Untuk Mengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Pembangunan Ekonomi,” vol. 17, no. 2, pp. 389–399, 2022.

- [12] V. No, S. Issue, N. Conference, P. A. Azis, R. Gustiawati, and F. Kurniawan, "Peran Fisiologi Olahraga Dalam Meningkatkan Kebugaran Fisik Dan Kesehatan Mental," vol. 23, no. 4, pp. 197–203, 2024.
- [13] A. A. Hamidah, I. Subagio, F. I. Olahraga, and U. N. Surabaya, "JPO: Jurnal Prestasi Olahraga," vol. 7, pp. 353–359, 2024.
- [14] P. Jambi, "Analisis Korelasi Pearson Jumlah Penduduk Dengan Jumlah Kendaraan Bermotor Di Provinsi Jambi," vol. 2, no. 1, pp. 39–44, 2023.
- [15] M. R. Kusnaldi, T. Gulo, and S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," vol. 3, no. 4, pp. 330–338, 2022, doi: 10.47065/Josyc.V3i4.2112.
- [16] K. Jember, "Jurnal Pendidikan MIPA," vol. 12, pp. 287–293, 2022.
- [17] P. P. Allorerung, A. Erna, and M. Bagussahrir, "Analisis Performa Normalisasi Data Untuk Klasifikasi K-Nearest Neighbor Pada Dataset Penyakit," vol. 9, no. 3, pp. 178–191, 2024.
- [18] A. Ardiansyah, J. Stanley, N. Tuandali, L. Sirri, and R. K. Hapsari, "Implementasi Algoritma K-Nearest Neighbor Dalam Prediksi Penyakit Jantung," vol. 7, no. 2, pp. 213–220, 2025.
- [19] G. Krishna and S. Yogeswara, "Pengklusteran Data Iris Menggunakan Metode Fuzzy," vol. 2, pp. 281–286, 2024.
- [20] M. Ardan, "Perbandingan Algoritma K-Means Dan Fuzzy C-Means Dalam Klasterisasi Data Pendidikan Di Wilayah," vol. 19, no. 1, pp. 73–82, 2024.
- [21] F. Novianti, Y. R. A. Yasmin, and D. C. R. Novitasari, "Penerapan Algoritma Fuzzy C-Means (FCM) Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Penyakit Menular Manusia," vol. 6, no. 1, pp. 23–33, 2022.
- [22] S. E. Saqila, I. P. Ferina, and A. Iskandar, "Analisis Perbandingan Kinerja Clustering Data Mining Untuk Normalisasi Dataset," vol. 5, 2023, doi: 10.30865/Json.V5i2.6919.
- [23] B. A. Manalu, M. Irwan, P. Nasution, U. Islam, and N. Sumatera, "Redudansi Data Pada Sistem Informasi Penjualan Analysis Of The Implementation Of Normalization And Reducing Data Redundancy In Sales Information Systems," vol. 4, no. 2, pp. 281–295, 2026.
- [24] W. D. Utami and I. A. Wijaya, "Optimasi Model Penugasan Berdasarkan Peramalan Layanan Backpropagation," vol. 6, no. 1, pp. 41–47, 2023.
- [25] N. A. Rahman and W. Yustanti, "Perbandingan Algoritma Klastering Dalam Pengelompokan Penjualan Produk Komputer," vol. 4, no. 4, pp. 95–100, 2023.
- [26] I. Maulana, W. Witanti, J. T. Informatika, U. Jenderal, A. Yani, and I. Artikel, "Klasterisasi Gaya Belajar Mahasiswa Berbasis VARK Dengan Algoritma DBSCAN untuk Personalisasi E-Learning," pp. 1619–1629, 2025, doi: 10.33364/Algoritma/V.22-2.2980.
- [27] S. Julianti and I. M. Widiartha, "Algoritma K-Means Untuk Clustering Provinsi Di Indonesia Berdasarkan Kasus Stunting," vol. 2, pp. 573–582, 2024.
- [28] Badan Pusat Statistik Indonesia, *Banyaknya Desa/Kelurahan Menurut Ketersediaan Lapangan Olahraga*, 2024, Mar. 10, 2026. Accessed: Jun. 13, 2026. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/OTc0IzI=/banyaknya-desa-kelurahan-menurut-ketersediaan-lapangan-olahraga--desa-.htm>