

KLASTERING WILAYAH DI JAWA TIMUR BERDASARKAN FAKTOR *UNMET NEED* MENGGUNAKAN FUZZY GUSTAFSON-KESSEL

Chrysilla Citra Windyadari¹, Aviolla Terza Damaliana², Mohammad Idhom³

Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur

Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur

e-mail: ¹21083010023@student.upnjatim.ac.id, ²aviolla.terza.sada@upnjatim.ac.id,
³idhom@upnjatim.ac.id

Abstract

The Family Planning Program is an effort to control the rate of population growth by regulating desired pregnancies. In its realization, the family planning program faces challenges in the form of unmet need (couples of childbearing age who do not use contraception). East Java Province in 2023 was recorded as the province with the third highest number of unmet need cases in Java. One method that can be used to analyze the phenomenon of unmet need is clustering analysis. Clustering analysis will help identify areas in East Java based on the priority level of the family planning program. Fuzzy Gustafson-Kessel (FGK) is one of the clustering methods developed as a refinement of the Fuzzy C-Means method. This study implements the Fuzzy Gustafson-Kessel (FGK) method with and without Principal Component Analysis (PCA) to cluster regions in East Java based on unmet need and determinant factors such as the availability of family planning facilities and resources. The results showed that the best model was obtained when using FGK with PCA, with the highest FSI value of 0.668 and XB of 0.235 at configuration $c = 4$ and $m = 3.5$. The clusters formed consist of 5 medium priority areas, 12 low priority areas, 9 high priority areas, and 12 developing priority areas. The results of this clustering can be used as a basis for policy makers in designing more effective intervention strategies to address unmet need in East Java.

Keyword: Clustering, Fuzzy Gustafson-Kessel, Fuzzy Silhouette Index, Xie-Beni, Unmet Need.

PENDAHULUAN

Dari tahun ke tahun jumlah penduduk di Indonesia mengalami peningkatan, salah satunya Provinsi Jawa Timur yang di tahun 2023 terjadi penambahan penduduk sebanyak 123.7 ribu jiwa (BPS Sulawesi Utara, 2023). Pertumbuhan penduduk yang tidak dapat terkendali dapat memberikan tekanan pada aspek kesejahteraan keluarga, keluarga berencana dan kependudukan. Pemerintah dalam upaya mengendalikan laju pertumbuhan penduduk menggagas Program Keluarga Berencana yang diterapkan secara serentak sejak tahun 1970. Program ini sebagai suatu upaya membentuk keluarga sejahtera dengan mengatur kehamilan yang diinginkan. Pada realisasinya program keluarga berencana menghadapi tantangan berupa kasus *unmet need*.

Unmet need merupakan persentase pasangan usia subur (PUS) yang tidak ingin hamil atau ingin menjarangkan kehamilan tetapi tidak menggunakan kontrasepsi (Safitri et al., 2021). Kondisi tersebut dapat meningkatkan risiko kehamilan yang tidak diinginkan sehingga mengakibatkan tingginya fertilitas di suatu wilayah. Pada tahun 2023, Provinsi Jawa Timur tercatat sebagai provinsi dengan jumlah kasus *unmet need* tertinggi ketiga di Pulau Jawa (BPS, 2024). Tingginya angka *unmet need* tidak hanya mengindikasikan rendahnya prevalensi penggunaan kontrasepsi, tetapi juga mencerminkan permasalahan mendalam dalam efektivitas layanan KB yang tersedia di wilayah tersebut. Salah satu metode yang dapat digunakan untuk menganalisis fenomena *unmet need* adalah analisis klastering. Analisis klastering merupakan teknik untuk mengelompokkan objek data ke dalam beberapa klaster, di mana setiap klaster berisi data dengan karakteristik serupa, sementara karakteristik antar klaster berbeda satu sama lain (Hidayati et al., 2022). Analisis klastering akan membantu mengidentifikasi wilayah-wilayah di Jawa Timur berdasarkan tingkat prioritas program KB.

Penelitian sebelumnya telah melakukan klastering wilayah berdasarkan kasus *unmet need*. Salah satunya adalah klastering kecamatan di Kabupaten Ponorogo ke dalam 4 klaster

berdasarkan jumlah pasangan usia subur peserta KB aktif dan *unmet need* tahun 2021 menggunakan metode K-Means (Bella Dwi Saputri & Diah Indriani, 2022). Dari hasil klastering tersebut diperoleh nilai WCSS sebesar 111.60 dan BCSS sebesar 1656.84. Penelitian lainnya memetakan 34 provinsi di Indonesia berdasarkan distribusi *unmet need* KB tahun 2015-2022 menggunakan metode K-Means, yang menghasilkan 3 klaster optimal dengan nilai Indeks Davies-Bouldin sebesar 1.248 (Samsir et al., 2023). Namun, penelitian terdahulu tersebut hanya melakukan klastering wilayah berdasarkan jumlah kasus *unmet need*. Memahami determinan *unmet need* di setiap wilayah penting untuk merancang intervensi yang tepat sasaran. Untuk itu, penelitian ini tidak hanya mengklaster berdasarkan *unmet need* tetapi juga memasukkan variabel determinan guna menghasilkan klaster yang lebih informatif.

Beberapa penelitian terdahulu telah mengidentifikasi faktor-faktor yang berkontribusi terhadap *unmet need*. Faktor-faktor seperti pemberian informasi oleh petugas kesehatan serta fasilitas pelayanan kesehatan memiliki hubungan signifikan terhadap keputusan penggunaan KB, khususnya IUD (Anggraini et al., 2022). Penelitian lainnya menunjukkan bahwa penggerak KB (PKB/PLKB, PPKBD, dan Sub PPKBD) berperan penting dalam meningkatkan partisipasi KB MKJP melalui promosi, penyuluhan, serta komunikasi, informasi, dan edukasi (KIE) kepada PUS, dengan nilai korelasi Pearson yang tinggi (Yuliati, 2021). Temuan ini menyoroti pentingnya variabel tenaga kesehatan, penyuluh KB, serta fasilitas pelayanan KB dalam meningkatkan cakupan kontrasepsi. Oleh karena itu, penelitian ini akan menganalisis klastering wilayah di Jawa Timur dengan fokus pada kasus *unmet need* dan faktor pemicu *unmet need* berupa variabel jumlah tempat KB, kampung KB, rata-rata ketersediaan alat kontrasepsi suntik 3 bulanan progestin, pil kombinasi, kondom, implan 2 batang, IUD, jumlah tenaga kesehatan, PKB/PLKB, serta PPKBD.

Fuzzy Gustafson-Kessel (FGK) merupakan metode klastering yang telah diterapkan dalam penelitian analisis klastering wilayah. Metode FGK dikembangkan sebagai penyempurnaan dari metode Fuzzy C-Means (FCM) dengan memanfaatkan jarak Mahalanobis. Penggunaan jarak tersebut membuat FGK tidak terbatas pada pola berbentuk bulat seperti FCM. Beberapa studi yang menunjukkan keunggulan FGK dibanding metode lainnya antara lain, klastering faktor kriminalitas di Indonesia yang membandingkan FCM dan FGK, dengan hasil bahwa FGK lebih unggul berdasarkan nilai S_w terkecil 0.83, S_B terbesar 0.15, dan rasio terkecil 5.53 (Destia & Kartikasari, 2023). Klasterisasi faktor risiko penularan TB di Jawa Timur, di mana FGK menghasilkan nilai Dunn Index tertinggi 0.974603, dibandingkan dengan FCM yang hanya sebesar 0.739133. Berdasarkan keunggulan tersebut, penelitian ini menggunakan Fuzzy Gustafson-Kessel sebagai metode analisis klastering wilayah.

Penelitian ini menggunakan 11 variabel analisis sehingga data berdimensi tinggi. Data berdimensi tinggi dapat memicu masalah rendahnya akurasi klasifikasi, kualitas klaster yang buruk, serta waktu komputasi yang lama (Hediyati & Suartana, 2021). Salah satu metode reduksi dimensi data yaitu Principal Component Analysis (PCA) banyak digunakan untuk mengoptimalkan kinerja pemodelan data. Penelitian terdahulu menunjukkan penerapan PCA sebelum klastering dapat meningkatkan kualitas klaster dengan nilai *silhouette coefficient* meningkat dari 0.61 menjadi 0.74 (Nurohmah et al., 2023). Namun, studi lain menunjukkan hasil sebaliknya, di mana klastering tanpa PCA menghasilkan nilai *silhouette* lebih tinggi sebesar 0.70 dibandingkan dengan PCA sebesar 0.67 (Wahyu & Rushendra, 2022). Oleh karena itu, penelitian ini akan membandingkan hasil klasterisasi FGK dengan dan tanpa PCA untuk mengevaluasi dampaknya terhadap hasil klastering *unmet need* dan determinannya di Provinsi Jawa Timur.

Tujuan dari membandingkan kedua pendekatan ini untuk menentukan metode yang lebih optimal dalam mengelompokkan wilayah berdasarkan kasus *unmet need* serta faktor-faktor determinannya. Evaluasi hasil klastering menggunakan *Fuzzy Silhouette Index* dan *Xie-Beni Index* untuk menilai kualitas klaster yang terbentuk. Dengan demikian, penelitian ini diharapkan dapat membantu pemerintah dalam merancang intervensi berbasis data guna menurunkan angka *unmet need* melalui peningkatan layanan KB yang tepat sasaran sesuai kebutuhan tiap wilayah.

METODE PENELITIAN

Dataset

Pengumpulan data merupakan proses memperoleh informasi sesuai dengan topik atau variabel yang telah ditetapkan (Fahrudin et al., 2022). Dalam penelitian ini data diperoleh dari sumber sekunder yaitu melalui website Sistem Informasi Keluarga (SIGA). Penelitian ini akan mengelompokkan 38 Kabupaten/Kota di Provinsi Jawa Timur berdasarkan 11 variabel yang diterangkan dalam Tabel 1.

Tabel 1. Variabel Penelitian

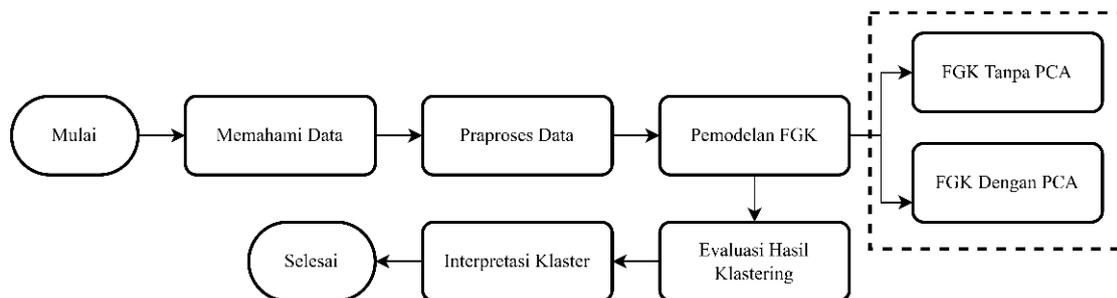
Variabel	Deskripsi
X ₁	Jumlah Kasus <i>Unmet need</i>
X ₂	Jumlah Tempat Pelayanan KB
X ₃	Jumlah Kampung KB
X ₄	Rata-Rata Kontrasepsi Suntik 3 Bulanan Progestin
X ₅	Rata-Rata Kontrasepsi Pil Kombinasi
X ₆	Rata-Rata Kontrasepsi Kondom
X ₇	Rata-Rata Kontrasepsi Implan 2 Batang
X ₈	Rata-Rata Kontrasepsi IUD
X ₉	Jumlah Tenaga Kesehatan
X ₁₀	Jumlah Penyuluh Keluarga Berencana (PKB)/ Jumlah Petugas Lapangan Keluarga Berencana (PLKB)
X ₁₁	Jumlah Pembantu Pembina Keluarga Berencana Desa (PPKBD)

Sampel dari dataset penelitian ditampilkan pada Tabel 2.

Tabel 2. Sampel Dataset Penelitian

Kabupaten/Kota	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁
Pacitan	7876	53	129	753.1	133.7	97.3	335.4	114.0	263	22	584
Ponorogo	18377	543	37	534.1	486.6	148.0	602.9	411.3	761	62	1024
Trenggalek	11495	236	86	1114.3	534.3	819.5	254.3	135.0	371	26	638
:	:	:	:	:	:	:	:	:	:	:	:
Kota Batu	4297	25	7	68.6	42.8	40.9	104.2	166.6	81	11	62

Prosedur Penelitian



Gambar 1. Tahapan Penelitian

Data yang telah dikumpulkan dalam penelitian ini akan diproses menggunakan *software* RStudio melalui serangkaian tahapan berikut.

1. Tahap awal dalam penelitian ini adalah memahami data yang digunakan. Pada langkah ini data akan dilakukan pengecekan *missing value*, data duplikat, serta statistika deskriptif. Pengecekan *missing value* bertujuan untuk menghindari kesalahan dalam analisis sementara pengecekan data duplikat bertujuan untuk memastikan ketepatan dan efisiensi dalam proses analisis data (Prasetya et al., 2025). Statistika deskriptif digunakan untuk memahami karakteristik data secara menyeluruh (Idhom et al., 2023).

2. Tahap praproses data dalam penelitian ini mencakup standarisasi data dan uji asumsi klastering. Standarisasi data bertujuan untuk mengubah skala data menjadi lebih seragam guna menghindari dominasi variabel tertentu dalam proses klastering misalnya saat perhitungan jarak (Fahrudin et al., 2021). Dalam penelitian ini, standarisasi dilakukan menggunakan metode Z-score dengan rumus berikut.

$$Z = \frac{x_i - \bar{x}}{s} \quad (1)$$

Dengan x_i sebagai data ke- i , \bar{x} sebagai rata-rata data, dan s sebagai standar deviasi. Z-score membantu mengetahui seberapa besar kemungkinan suatu nilai muncul dalam distribusi normal dan membandingkan posisi relatif dua nilai dari distribusi yang berbeda (Hindrayani et al., 2020). Selanjutnya, dilakukan uji asumsi klastering kecukupan sampel dan nonmultikolinearitas untuk memastikan validitas data sebelum klastering. Uji kecukupan sampel menggunakan Kaiser Mayer Olkin (KMO). Jika nilai KMO > 0.5 maka asumsi kecukupan sampel terpenuhi. Rumus uji KMO sebagai berikut (Mahmudan, 2020).

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (2)$$

Dengan p sebagai jumlah variabel, r_{ij} sebagai koefisien korelasi antara variabel i dan j , serta a_{ij} sebagai koefisien korelasi parsial antara variabel i dan j . Uji nonmultikolinearitas dilakukan menggunakan VIF (Variance Inflation Factor). Nilai VIF < 10 menunjukkan tidak terdapat multikolinearitas. Adapun rumus dari VIF sebagai berikut (Mahmudan, 2020).

$$VIF = \frac{1}{1 - R_{yx_1x_2}^2} \quad (3)$$

Di mana $R_{yx_1x_2}^2$ merupakan koefisien determinasi antara y dengan variabel bebas lainnya.

3. Implementasi klastering menggunakan model Fuzzy Gustafson-Kessel (FGK) dengan dua pendekatan yaitu menerapkan PCA dan tanpa PCA. Konfigurasi parameter yang digunakan untuk kedua pendekatan mencakup jumlah kluster (c) = 2, 3, 4 dan pangkat pembobot (m) = 2, 2.5, 3, 3.5, 4.
4. Evaluasi hasil klastering, pemodelan FGK dengan dan tanpa PCA dievaluasi menggunakan *Fuzzy Silhouette Index* (FSI) dan *Xie-Beni Index* (XB) untuk menilai kualitas kluster yang terbentuk. Kluster optimal dipilih berdasarkan kriteria FSI tertinggi dengan nilai XB yang lebih rendah dari FSI ($XB < FSI$).
5. Terakhir, interpretasi hasil klastering dari model klastering terbaik.

Principal Component Analysis (PCA)

Proses analisis PCA dijelaskan dalam tahapan berikut (Hediyati & Suartana, 2021).

1. Menentukan nilai rata-rata setiap dimensi menggunakan persamaan berikut.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4)$$

2. Menentukan matriks kovarians (C_X) untuk menganalisis korelasi setiap dimensi data.

$$C_X = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (5)$$

3. Menentukan *eigenvector* (v_m) dan *eigenvalue* (λ_m) dari matriks kovarians.

$$C_X v_m = \lambda_m v_m \quad (6)$$

4. Memilih komponen utama berdasarkan *eigenvector* dengan *eigenvalue* terbesar, diurutkan dari nilai terbesar ke terkecil.
5. PCA menghasilkan dataset baru yang terdiri dari komponen utama, di mana informasi yang lebih signifikan dari dataset asli tetap terjaga.

Pemilihan jumlah komponen didasarkan pada persentase kumulatif varians yang mampu menjelaskan total keragaman variabel asal (Pramana et al., 2023). Persentase kumulatif yang dianggap cukup sebesar 70% hingga 80% (Oktarina et al., 2024).

Fuzzy Gustafson-Kessel

Fuzzy Gustafson-Kessel (FGK) merupakan pengembangan lebih lanjut dari metode Fuzzy C-Means (FCM). FGK selaras dengan prinsip fuzzy, mampu mengelompokkan data ke dalam suatu kluster melalui nilai derajat keanggotaan. Derajat keanggotaan memiliki nilai diantara 0 hingga 1 (Ragestu & Sibarani, 2020). Berbeda dengan FCM yang mengandalkan jarak Euclidean, FGK menggunakan jarak Mahalanobis sebagai metrik jarak adaptif. Metrik ini diperbarui pada setiap iterasi berdasarkan matriks kovarians fuzzy yang dihitung dari data input (Paradilla et al., 2023). Pendekatan ini memungkinkan identifikasi kluster dengan bentuk yang lebih fleksibel (Nurzida et al., 2024). Berikut tahapan metode Fuzzy Gustafson-Kessel (Paradilla et al., 2023).

1. Menambahkan data sebagai X (matriks $n \times p$), n adalah jumlah data input dan p adalah jumlah variabel penelitian.
2. Menetapkan jumlah kluster ($c \geq 2$), pangkat pembobotan ($m > 1$), maksimum iterasi (t_{max}), *expected error* (ϵ), dan inisialisasi awal fungsi objektif ($P_0 = 0$).
3. Membangkitkan nilai acak u_{ik} , di mana i merupakan indeks data ke- i dengan jumlah data sebanyak n ($i = 1, 2, 3, \dots, n$) dan k sebagai indeks kluster dengan rentang $1 \leq k \leq c$. Nilai u_{ik} sebagai inisialisasi awal elemen-elemen dari matriks partisi U_0 .

$$U_0 = \begin{bmatrix} u_{1,1} & u_{1,2} & \dots & u_{1,c} \\ u_{2,1} & u_{2,2} & \dots & u_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n,1} & u_{n,2} & \dots & u_{n,c} \end{bmatrix} \text{ dengan } \sum_{k=1}^c u_{ik} = 1$$

4. Menghitung pusat kluster ke- k (v_{kj})

$$v_{kj} = \frac{\sum_{i=1}^n (u_{ik}^m \times x_{ij})}{\sum_{i=1}^n u_{ik}^m} \quad (7)$$

5. Menentukan matriks kovarian kluster (F_k)

$$F_k = \frac{1}{\sum_{i=1}^n u_{ik}^m} \times R_k^T \times R_k^* \quad (8)$$

Dengan,

$$R_k = \begin{bmatrix} (x_{11} - v_{k1}) & (x_{12} - v_{k2}) & \dots & (x_{1p} - v_{kp}) \\ (x_{21} - v_{k1}) & (x_{22} - v_{k2}) & \dots & (x_{2p} - v_{kp}) \\ \vdots & \vdots & \ddots & \vdots \\ (x_{n1} - v_{k1}) & (x_{n2} - v_{k2}) & \dots & (x_{np} - v_{kp}) \end{bmatrix}$$

$$R_k^* = \begin{bmatrix} u_{1k}^m \times (x_{11} - v_{k1}) & u_{1k}^m \times (x_{12} - v_{k2}) & \dots & u_{1k}^m \times (x_{1p} - v_{kp}) \\ u_{2k}^m \times (x_{21} - v_{k1}) & u_{2k}^m \times (x_{22} - v_{k2}) & \dots & u_{2k}^m \times (x_{2p} - v_{kp}) \\ \vdots & \vdots & \ddots & \vdots \\ u_{nk}^m \times (x_{n1} - v_{k1}) & u_{nk}^m \times (x_{n2} - v_{k2}) & \dots & u_{nk}^m \times (x_{np} - v_{kp}) \end{bmatrix}$$

6. Menghitung jarak Mahalanobis (D_{ik}^2)

$$D_{ik}^2 = (R_k \times A_k)(R_k^T) \quad (9)$$

A_k sebagai fungsi jarak adaptif (*adaptive distance norm*) dengan perhitungan berikut.

$$A_k = \left[\left(\det(F_k)^{\frac{1}{n}} \right) \right] (F_k^{-1}) \quad (10)$$

F_k^{-1} merupakan inverse matriks dari kovarian *fuzzy*.

$$D_k^2 = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \text{ untuk } S_{ik}^2 = \begin{bmatrix} d_{11} \\ d_{22} \\ \vdots \\ d_{nn} \end{bmatrix} = \begin{bmatrix} S_{1k} \\ S_{2k} \\ \vdots \\ S_{nk} \end{bmatrix}$$

Matriks D_k^2 berbentuk persegi, di mana nilai jarak Mahalanobis diperoleh dari elemen-elemen pada diagonal utamanya yang dinotasikan sebagai S_{ik}^2 .

7. Menentukan fungsi objektif pada iterasi ke- t (P_t)

$$P_t = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m \times S_{ik}^2 \quad (11)$$

8. Memperbarui keanggotaan matriks u_{ik} ($u_{ik(new)}$)

$$u_{ik} = \left[\sum_{q=1}^c \left(\frac{S_{ik}}{S_{iq}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (12)$$

S_{iq} sebagai jarak Mahalanobis antara titik data i dengan kluster q .

9. Mengulangi langkah ke-4 hingga ke-8 hingga salah satu dari dari kondisi berikut terpenuhi: selisih fungsi objektif $(|P_t - P_{t-1}|) < \varepsilon$; iterasi telah mencapai batas maksimal $(t > t_{max})$.

Evaluasi Hasil Klustering

Evaluasi hasil klustering yang akan digunakan untuk menentukan jumlah kluster yang optimal dalam model klustering fuzzy adalah *Fuzzy Silhouette Index* (FSI) dan *Xie-Beni* (XB). FSI memiliki rentang nilai -1 hingga 1, yang berarti semakin tinggi nilai FSI menunjukkan hasil kluster yang baik (Fahmi Ahmad Aulia et al., 2022). FSI dirumuskan sebagai berikut.

$$s_k = \frac{b_{ik} - a_{jk}}{\max(a_{ik}, b_{jk})} \quad (13)$$

$$FS = \frac{\sum_{i,j=1}^n (u_{ik} - u_{jk})^\alpha s_k}{\sum_{i,j=1}^n (u_{ik} - u_{jk})^\alpha} \quad (14)$$

Di mana s_k sebagai *silhouette coefficient*, b_{ij} sebagai rata-rata jarak suatu titik data j ke titik-titik dalam kluster terdekat (antar-kluster minimal), a_{pj} sebagai rata-rata jarak suatu titik data i ke seluruh titik dalam klasternya sendiri (intra-kluster), u_{pj} sebagai nilai derajat keanggotaan terbesar pertama kolom matriks ke- k , u_{qj} sebagai nilai derajat keanggotaan terbesar kedua kolom matriks ke- k , dan α sebagai bobot koefisien ($\alpha \geq 0$).

Xie-Beni Index (XB), salah satu indeks validitas yang mengukur rasio antara total variasi dalam kluster (S) dan jarak pemisah antar kluster (d_{min}) (Putri & Rochmawati, 2021). Nilai XB yang minimal menunjukkan hasil kluster yang lebih baik.

$$XB(c) = \frac{S}{d_{min}} \quad (15)$$

$$S = \frac{\sum_{i=1}^n \sum_{k=1}^c (u_{ik})^m d_{ik}^2(x_i, v_k)}{n} \quad (16)$$

$$d_{min} = \min_{i \neq k} d_{ik}^2(v_i, v_k) \quad (17)$$

Dengan $d_{ik}(x_i, v_k)$ merupakan jarak data ke pusat kluster dan $d_{ik}(v_i, v_k)$ adalah jarak antar pusat kluster. Model kluster yang optimal dalam penelitian ini dipilih berdasarkan nilai FSI tertinggi dan nilai XB yang lebih rendah dari FSI ($XB < FS$) (Fahmi Ahmad Aulia et al., 2022). Pendekatan ini digunakan karena lebih mencerminkan pemisahan kluster yang baik serta mempertimbangkan validitas internal kluster.

HASIL DAN PEMBAHASAN

1. Analisis Statistika Deskriptif

Pada tahap memahami data, tidak ditemukan *missing value* maupun data duplikat. Selanjutnya, dilakukan analisis statistik deskriptif untuk memberikan gambaran umum mengenai data penelitian. Berdasarkan hasil analisis statistik deskriptif Gambar 2, terdapat perbedaan yang cukup besar dalam distribusi beberapa variabel penelitian.

```
> describe(data[2:12])
vars  n    mean    sd    min    max
Unmet_Need      1 38 17192.32 14342.45 2124.00 72201.00
Tempat_KB       2 38  167.34   127.98    24.00   543.00
Kampung_KB      3 38  135.71   107.47     4.00  466.00
SUNTIK_3_BULANAN_PROGESTIN 4 38 1788.86 1855.79  43.00  8367.00
PIL_KOMBINASI   5 38 2143.18 3357.87  42.82 15466.58
KONDOM          6 38  280.75  215.89  32.92  848.75
IMPLAN_2_BATANG 7 38  393.18  301.43  30.25 1350.33
IUD             8 38  214.59  167.51   9.58  633.33
Tenaga_Kesehatan 9 38  420.00  205.23  81.00  766.00
PKB_PLKB       10 38   44.13   25.12  10.00  105.00
kader_PPKBD    11 38  788.89  499.81  35.00 1847.00
```

Gambar 2. Statistika Deskriptif

Variabel jumlah kasus *unmet need* menunjukkan adanya ketimpangan yang signifikan. Rata-rata jumlah kasus *unmet need* di berbagai daerah adalah 17,192 kasus dengan standar deviasi sebesar 14,342 kasus, artinya terdapat daerah dengan jumlah kasus yang sangat tinggi maupun sangat rendah. Hal ini diperkuat dengan rentang nilai minimal-maksimal kasus *unmet need* yaitu 2,124 hingga 72,201 kasus. Selanjutnya, variabel jumlah tempat pelayanan KB dan kampung KB menunjukkan variasi antar daerah. Rata-rata tempat pelayanan KB sebesar 167 unit (standar deviasi = 127) dan kampung KB sebesar 135 unit (standar deviasi = 107) yang mencerminkan ketimpangan ketersediaan fasilitas KB antar wilayah. Pada variabel ketersediaan alat kontrasepsi, hasil analisis deskriptif menunjukkan bahwa jumlah ketersediaan beberapa jenis kontrasepsi sangat bervariasi antar daerah. Selain itu, tenaga kesehatan dan petugas KB juga menunjukkan distribusi yang tidak merata per daerah.

2. Praproses Data

Data terlebih dahulu distandarisasi menggunakan Z-score, kemudian dilakukan uji kecukupan sampel menggunakan KMO dan nonmultikolinearitas menggunakan VIF. Berikut uji kecukupan sampel melalui perhitungan nilai KMO.

Hipotesis uji:

H_0 : Ukuran data memadai untuk dilakukan analisis klustering

H_1 : Ukuran data tidak memadai untuk dilakukan analisis klustering

Statistik uji:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} = \frac{17.7985}{17.7985 + 6.6459} = 0.7281$$

Keputusan: Nilai KMO sebesar 0.7281 berada di atas batas minimal 0.5. Maka, H_0 diterima sehingga data dinyatakan layak untuk dilakukan analisis klustering lebih lanjut.

Selanjutnya, dilakukan uji nonmultikolinearitas menggunakan VIF. Untuk mendukung hasil pengujian ini, disajikan satu contoh perhitungan VIF. Dalam hal ini, variabel *unmet need* diregresikan terhadap seluruh variabel independen lainnya untuk memperoleh nilai koefisien determinasi (R^2), yang digunakan dalam perhitungan VIF.

Hipotesis uji:

H_0 : Tidak terdapat hubungan linier yang signifikan antar variabel

H_1 : Terdapat hubungan linier yang signifikan antar variabel

Statistik uji:

Tabel 3. Hasil Perhitungan VIF untuk Variabel *Unmet Need*

Variabel	Koefisien Determinasi (R^2)	VIF
Tempat KB	0.5657	2.3027
Kampung KB	0.4806	1.9251
Suntik 3 Bulanan Progestin	0.4252	1.7399
Pil Kombinasi	0.4495	1.8165
Kondom	0.4062	1.6842
Implan 2 Batang	0.7009	3.3436
IUD	0.6661	2.9947
Tenaga Kesehatan	0.7525	4.0407
PKB/PLKB	0.5412	2.1798
Kader PPKBD	0.7393	3.8351

Keputusan: Seluruh nilai VIF berada di bawah ambang batas 10, sehingga tidak terdapat hubungan linier yang signifikan antar variabel. Dengan demikian, H_0 diterima, dan data dapat digunakan untuk analisis klustering.

3. Pemodelan Fuzzy Gustafson-Kessel (FGK)

Dalam penelitian ini, dilakukan percobaan pemodelan dengan variasi jumlah kluster

(c) dan pangkat pembobot (m). Secara umum, nilai pangkat pembobot yang digunakan adalah $m = 2$. Menurut K.-L. Wu, rentang nilai pangkat pembobot yang baik berada dalam interval $m \in [1.5, 4]$ (Tambuwun et al., 2020). Berdasarkan hal tersebut, penelitian ini menggunakan konfigurasi parameter $m \in [2, 4]$ dengan kelipatan 0.5. Konfigurasi parameter diterapkan untuk pemodelan FGK dengan dan tanpa PCA. Rincian konfigurasi parameter disajikan dalam Tabel 4.

Tabel 4. Konfigurasi Parameter

Jumlah Kluster (c)	Pangkat Pembobot (m)	Iterasi Maksimum (t)	Expected Error (ϵ)
2,3,4	2,2.5,3,3.5,4	1000	0.00001

Pendekatan FGK dengan PCA memerlukan tahap awal menentukan jumlah komponen berdasarkan ambang minimal persentase kumulatif varians 70%. Berdasarkan Tabel 5, dipilih 4 komponen utama yang secara kumulatif menjelaskan 77.6% keragaman data sehingga dianggap representatif untuk proses klusterisasi.

Tabel 5. Analisis Komponen

Komponen	Eigenvalue	Persentase Varians	Persentase Kumulatif Varians
1	4.718	42.9	42.9
2	1.842	16.7	59.6
3	1.079	9.8	69.4
4	0.897	8.2	77.6
5	0.764	6.9	84.5
6	0.487	4.4	89.0
7	0.397	3.6	92.6
8	0.341	3.1	95.7
9	0.220	2.0	97.7
10	0.156	1.4	99.1
11	0.099	0.9	100

Pendekatan tanpa PCA menggunakan seluruh variabel asli secara langsung dalam proses klustering, sementara pendekatan dengan PCA menerapkan reduksi dimensi untuk mengekstrak 4 komponen utama sebagai input dalam pemodelan FGK. Berdasarkan hasil analisis yang disajikan dalam Tabel 6, klaster dikatakan optimal jika memiliki nilai FSI tertinggi dan nilai XB yang lebih rendah dari FSI. Secara umum, hasil penelitian menunjukkan bahwa penerapan PCA pada model FGK dapat meningkatkan hasil evaluasi klaster dibandingkan model FGK tanpa PCA.

Tabel 6. Hasil Evaluasi Klaster Model FGK Dengan dan Tanpa PCA

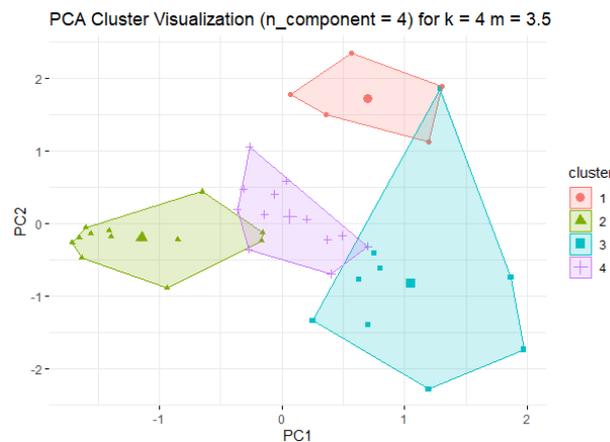
Jumlah Kluster (c)	Pangkat Pembobot (m)	Fuzzy Silhouette Index (FSI)	Xie-Beni Index (XB)	Keterangan
2	2	0.178	1.844	Tanpa PCA
	2.5	0.127	1.068	
	3	0.135	0.673	
	3.5	0.201	0.407	
	4	0.327	0.273	
2	2	0.362	0.552	PCA
	2.5	0.382	0.390	
	3	0.508	0.246	
	3.5	0.522	0.171	
	4	0.538	0.124	
3	2	-0.012	5.859	Tanpa PCA
	2.5	0.076	3.892	
	3	0.054	4.684	
	3.5	0.236	0.170	
	4	0.254	0.091	
3	2	0.272	0.571	PCA
	2.5	0.519	0.258	
	3	0.586	0.270	
	3.5	0.695	0.896	
	4	0.833	23409830	

Tabel 6. Hasil Evaluasi Kluster Model FGK Dengan dan Tanpa PCA

Jumlah Kluster (<i>c</i>)	Pangkat Pembobot (<i>m</i>)	Fuzzy Silhouette Index (FSI)	Xie-Beni Index (XB)	Keterangan
4	2	-0.119	6.252	Tanpa PCA
	2.5	0.135	3.324	
	3	0.023	7.022	
	3.5	-0.052	5.284	
	4	0.123	0.063	
4	2	0.149	1.108	PCA
	2.5	0.467	0.226	
	3	0.628	0.162	
	3.5	0.668	0.235	
	4	0.836	8024690000	

Nilai FSI tertinggi diperoleh saat menerapkan PCA dalam model FGK. Parameter $c = 4$ dan $m = 4$ menghasilkan nilai FSI tertinggi sebesar 0.836, tetapi nilai XB yang lebih tinggi dari FSI menunjukkan kluster yang terbentuk belum optimal. Kondisi serupa terjadi pada parameter $c = 3$ untuk $m = 4$ dan $m = 3.5$ dengan FSI sebesar 0.833 serta 0.695, di mana keduanya memiliki nilai XB lebih tinggi dari FSI sehingga tidak memenuhi kriteria kluster yang optimal. Sebaliknya, pada parameter $c = 4$ dan $m = 3.5$, diperoleh FSI sebesar 0.668 dengan nilai XB lebih rendah, yaitu 0.235. Konfigurasi ini menunjukkan kluster yang lebih optimal karena memenuhi kriteria $XB < FSI$, mengindikasikan bahwa data dalam setiap kluster lebih terkelompok dengan baik.

Hasil visualisasi Gambar 3 memperlihatkan pemisahan kluster yang lebih jelas dan heterogen, dengan tumpang tindih antar kluster yang lebih minimal. Dengan demikian, model terbaik diperoleh dengan menerapkan PCA pada parameter $c = 4$ dan $m = 3.5$.



Gambar 3. Visualisasi Hasil Kluster $c = 4$ dan $m = 3.5$

4. Interpretasi Kluster

Tabel 7 menyajikan jumlah anggota dalam masing-masing kluster berdasarkan pemodelan terbaik Fuzzy Gustafson-Kessel (FGK) dengan PCA.

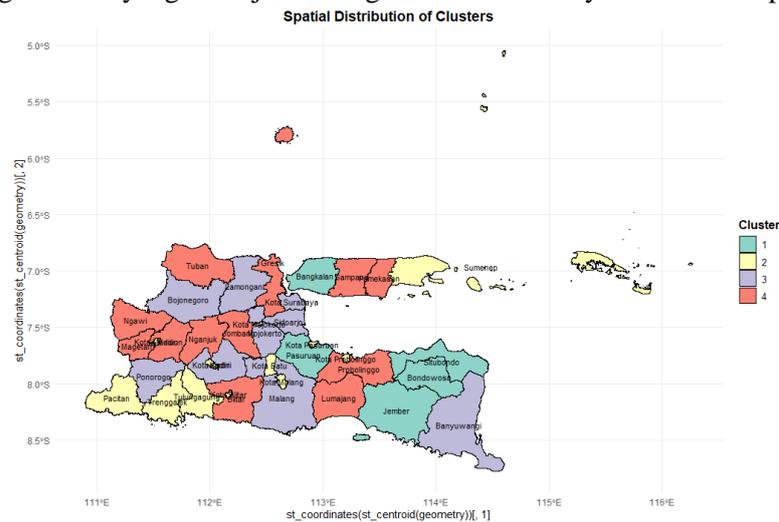
Tabel 7. Anggota Kluster

Kluster	Anggota Kluster	Jumlah Anggota Kluster
1	Jember, Bondowoso, Situbondo, Pasuruan, Bangkalan	5
2	Pacitan, Trenggalek, Tulungagung, Sumenep, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu	12
3	Ponorogo, Kediri, Malang, Banyuwangi, Sidoarjo, Mojokerto, Bojonegoro, Lamongan, Kota Surabaya	9

Tabel 7. Anggota Klaster

Klaster	Anggota Klaster	Jumlah Anggota Klaster
4	Blitar, Lumajang, Probolinggo, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Tuban, Gresik, Sampang, Pamekasan	12

Berdasarkan hasil klustering, klaster pertama terdiri dari 5 wilayah, klaster kedua dan keempat masing-masing memiliki 12 wilayah, sedangkan klaster ketiga mencakup 9 wilayah. Untuk memahami distribusi klaster, hasil klasterisasi divisualisasikan dalam bentuk peta guna memberikan gambaran yang lebih jelas mengenai sebaran wilayah dalam setiap klaster.



Gambar 4. Peta Visualisasi Klaster

Interpretasi hasil klustering diperoleh dari nilai rata-rata variabel untuk setiap klaster. Variabel X_1 merepresentasikan kasus *unmet need*, $X_2 - X_3$ sebagai fasilitas KB, $X_4 - X_8$ sebagai ketersediaan alat kontrasepsi, X_9 sebagai tenaga kesehatan, dan $X_{10} - X_{11}$ sebagai penggerak KB.

Tabel 8. Nilai Rata-Rata Variabel Tiap Klaster

Klaster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	16902.6	257.2	152.6	5290	8665.8	386.8	530.3	110.6	510	60.2	851.2
2	8703.4	61.6	52.7	683.4	526	196	147.6	116	198.2	18.9	322.6
3	31791.1	229.2	197	1705.7	1880.9	437.9	696.3	433.7	610	52.2	1201.4
4	14852.8	189.3	165.8	1498	1239.3	203.4	354.3	192.2	461.8	56.6	919.8

- Klaster 1 menunjukkan jumlah *unmet need* cukup tinggi kedua. Fasilitas KB, ketersediaan alat kontrasepsi, serta jumlah tenaga kesehatan dan penggerak KB tergolong memadai. Wilayah ini termasuk prioritas menengah, dengan kebutuhan KB besar dan dukungan infrastruktur yang cukup, namun masih diperlukan peningkatan efektivitas program untuk menekan *unmet need*.
- Klaster 2 memiliki *unmet need* terendah, dengan jumlah fasilitas KB, tenaga kesehatan, dan penggerak KB paling sedikit. Ketersediaan alat kontrasepsi juga rendah. Wilayah ini termasuk kategori prioritas rendah karena kebutuhan dan layanan KB sudah seimbang.
- Klaster 3 menunjukkan *unmet need* tertinggi, meskipun layanan KB cukup memadai. Tingkat penggunaan kontrasepsi yang rendah menjadi tantangan, sehingga wilayah ini dikategorikan sebagai prioritas tinggi.
- Klaster 4 mencatat *unmet need* tertinggi ketiga, dengan fasilitas dan dukungan KB tergolong menengah-tinggi. Kebutuhan dan layanan KB cenderung seimbang, sehingga masuk dalam kategori prioritas berkembang.

KESIMPULAN

Berdasarkan hasil klastering menggunakan Fuzzy Gustafson-Kessel, diperoleh hasil terbaik saat menerapkan PCA pada model. Konfigurasi optimal diperoleh untuk $c = 4$ dan $m = 3.5$, dengan nilai FSI tertinggi sebesar 0.668 serta XB sebesar 0.235. Klaster 1 meliputi 5 wilayah dengan kategori prioritas menengah, klaster 2 meliputi 12 wilayah dengan kategori prioritas rendah, klaster 3 meliputi 9 wilayah dengan kategori prioritas tinggi, dan klaster 4 meliputi 12 wilayah dengan kategori prioritas berkembang. Hasil klasterisasi ini menunjukkan adanya variasi dalam kebutuhan KB, ketersediaan fasilitas, dan sumber daya penggerak KB.

SARAN

Penelitian selanjutnya dapat menambahkan variabel lain seperti dari aspek sosial untuk memperdalam faktor-faktor yang mempengaruhi *unmet need* serta dapat melakukan eksplorasi pemodelan klaster lainnya.

DAFTAR PUSTAKA

- Anggraini, C., Putri, R., & Rini, A. S. (2022). Hubungan fasilitas pelayanan kesehatan, sumber informasi dan persepsi ibu terhadap penggunaan kontrasepsi IUD. *SENTRI: Jurnal Riset Ilmiah*, 1(3), 685–694.
- Bella Dwi Saputri, & Diah Indriani. (2022). Pemetaan Cakupan Pengguna KB Aktif Dan *Unmet Need* Di Kabupaten Ponorogo Tahun 2021. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(1). <https://doi.org/10.36456/jstat.vol15.no1.a5446>
- BPS. (2024, January 2). *Persentase Unmet Need Pelayanan Kesehatan Menurut Provinsi (Persen), 2021-2023*. Badan Pusat Statistika. <https://www.bps.go.id/id/statistics-table/2/MTQwMiMy/unmet-need-pelayanan-kesehatan-menurut-provinsi.html>
- BPS Sulawesi Utara. (2023). *Jumlah Penduduk Menurut Provinsi di Indonesia (Ribu Jiwa), 2021-2023*. Badan Pusat Statistika Sulawesi Utara. <https://sulut.bps.go.id/id/statistics-table/2/OTU4IzI=/jumlah-penduduk-menurut-provinsi-di-indonesia.html>
- Destia, B., & Kartikasari, M. D. (2023). Comparison of Fuzzy C-Means and Fuzzy Gustafson-Kessel Clustering Methods in Provincial Grouping in Indonesia Based on Criminality-Related Factors. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17(2), 1093–1102.
- Fahmi Ahmad Aulia, Irandia Ginanjar, & Achmad Bachrudin. (2022). Pengelompokan Desa Di Kabupaten Bandung Berdasarkan Indikator Pembangunan Desa Tahun 2021 Dengan Fuzzy C-Medoids Clustering. *Seminar Nasional Statistika Aktuaria I (2022)*, 1(1), 260–268.
- Fahrudin, T. M., Riyantoko, P. A., Hindrayani, K. M., & Swari, M. H. P. (2021). Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators Using Agglomerative Hierarchical Clustering. *Telematika*, 18(2), 194–210. <https://doi.org/10.31315/telematika.v18i2.4786>
- Fahrudin, T. M., Sari, A. R. F., Lisanthoni, A., & Lestari, A. A. D. (2022). Analisis Speech-To-Text Pada Video Mengandung Kata Kasar Dan Ujaran Kebencian Dalam Ceramah Agama Islam Menggunakan Interpretasi Audiens Dan Visualisasi Word Cloud. *SKANIKA: Sistem Komputer Dan Teknik Informatika*, 5(2), 190–202.
- Hediyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro. *JIEET (Journal of Information Engineering and Educational Technology)*, 5(2), 49–54.
- Hidayati, S., Darmaliana, A. T., & Riski, R. (2022). Comparison of K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN for Village Grouping in Surabaya Based on Poverty Indicators. *Jurnal Pendidikan Matematika (Kudus)*, 5(2), 185–198.

- Hindrayani, K. M., Fahrudin, T. M., Aji, R. P., & Safitri, E. M. (2020). Indonesian Stock Price Prediction Including Covid19 Era Using Decision Tree Regression. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 344–347.
- Idhom, M., Fauzi, A., Trimono, T., & Riyantoko, P. (2023). Time Series Regression: Prediction of Electricity Consumption Based on Number of Consumers at National Electricity Supply Company. *TEM Journal*, 12(3), 1575.
- Mahmudan, A. (2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering. *J. Mat. Stat. Dan Komputasi*, 17(1), 1–13.
- Nurohmah, Y., Mayasari, R., & Sari, B. N. (2023). Optimalisasi Performa K-Means Clustering Dengan Pca Dalam Analisis Tingkat Kemiskinan Di Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1657–1665.
- Nurzida, A., Utami, I. T., & Rochayani, M. Y. (2024). Perbandingan Metode Fuzzy C-Means Dan Gustafson-Kessel Dalam Penentuan Cluster Tingkat Risiko Penularan Tuberculosis Terhadap Penyakit Di Jawa Timur. *Jurnal Gaussian*, 13(2), 373–382.
- Oktarina, C. R., Sriliana, I., Sidik, E. N. F., & Firmansyah, M. A. (2024). Pemodelan Data Geospasial Balita Kurang Gizi Dengan Pendekatan Geographically Weighted Regression Principal Component Analysis. *Jurnal Gaussian*, 13(2), 339–350.
- Paradilla, Y. S., Hayati, M. N., & Sifriyani, S. (2023). Implementation of The Fuzzy Gustafson-Kessel Method on Grouping Districts/Cities in Kalimantan Island Based on Poverty Issues Factors. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17(1), 125–134.
- Pramana, S., Yuniarto, B., Santoso, I., Nooraeni, R., & Suadaa, L. H. (2023). *Data Mining dengan R, Konsep dan Implementasi*. In Media.
- Prasetya, D. A., Sari, A. P., Idhom, M., & Lisanthoni, A. (2025). Optimizing Clustering Analysis to Identify High-Potential Markets for Indonesian Tuber Exports. *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(1), 113–122.
- Putri, F. A. A., & Rochmawati, N. (2021). Penerapan Algoritma Gustafson-Kessel untuk Clustering Tweets Mention Akun Go-jek dan Grab Indonesia. *Journal of Informatics and Computer Science (JINACS)*, 2(03), 204–212.
- Ragestu, F. D., & Sibarani, A. J. P. (2020). Penerapan Metode Fuzzy Tsukamoto Dalam Pemilihan Siswa Teladan di Sekolah. *Teknika*, 9(1). <https://doi.org/10.34148/teknika.v9i1.251>
- Safitri, H., Siregar, K. N., Eryando, T., Herdayati, M., Rahmadewi, R., & Irawaty, D. K. (2021). Pemberian Layanan Keluarga Berencana Berpengaruh Penting Terhadap Kejadian *Unmet Need*: Analisis Lanjut Data SDKI 2017. *Jurnal Biostatistik, Kependudukan, Dan Informatika Kesehatan*, 1(2), 66–78. <https://doi.org/10.51181/bikfokes.v1i2.4751>
- Samsir, S., Watrionthos, R., & Suryadi, S. (2023). Distribusi Spasial *Unmet Need* Pelayanan Kesehatan dengan Algoritma K-Means untuk Pemetaan Provinsi di Indonesia. *Bulletin of Information Technology (BIT)*, 4(3), 362–368.
- Tambuwun, C. H., Langi, Y. A. R., & Rindengan, A. J. (2020). Estimasi Bobot Parameter M Pada Fuzzy C-Means Menggunakan Analisis Robust Dengan Simulasi Data Spasial. *D'Cartesian: Jurnal Matematika Dan Aplikasi*, 9(1), 50–55.
- Wahyu, A., & Rushendra, R. (2022). Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 8(1), 174–179.
- Yuliati, I. F. (2021). Peramalan Dan Analisis Hubungan Faktor Penggerakan Lini Lapangan Dalam Meningkatkan Peserta Kb Aktif Mkj. *Jurnal Keluarga Berencana*, 6(2), 35–48.