

Optimalisasi Algoritma Fuzzy Gustafson-Kessel dengan Variasi Parameter Weighting Exponent untuk Clustering Data Risiko Depresi Pelajar

Amira Adelia Putri,
Intan Nisfulaila*, and
Erna Herawati

Program Studi Matematika, Fakultas
Sains dan Teknologi,
Universitas Islam Negeri Maulana Malik
Ibrahim Malang,
Indonesia

Article History

Received 25 Mei 2026
Revised 6 Juni 2026
Accepted 28 Juni 2026
Published 30 Juni 2026



Copyright © 2026 by Authors, Published by
JRMM Group. This is an open access
article under the CC BY-SA License.



Abstract. This study aims to analyse the impact of variations in the weighting exponent on the Fuzzy Gustafson-Kessel algorithm in the clustering process of depression risk data among students. The data used comprised 502 respondents with seven variables representing academic, financial, and lifestyle factors. The research stages included data preprocessing using label encoding and normalisation, followed by clustering on several variations of the weighting exponent value. The quality of the clustering results was evaluated using the Xie-Beni index as a measure of the compactness and separation of the clusters. The results showed that smaller values of the weighting exponent produced higher quality clusters, whereas increasing the exponent value caused the index value to rise. The best combination was obtained with three clusters and a weighting exponent value of 1.1, which produced the lowest Xie-Beni index value. The clustering results were able to group the data into three categories of depression risk, namely low, moderate, and high. These findings demonstrate that adjusting the weighting exponent parameter in the Fuzzy Gustafson-Kessel algorithm plays a crucial role in improving clustering quality.

Keywords: Fuzzy Gustafson-Kessel; Weighting Exponent; Xie-Beni Index; Clustering; Depression

Abstrak. Penelitian ini bertujuan untuk menganalisis pengaruh variasi nilai *weighting exponent* pada algoritma *Fuzzy Gustafson-Kessel* dalam proses *clustering* data risiko depresi pelajar. Data yang digunakan berjumlah 502 responden dengan tujuh variabel yang merepresentasikan faktor akademik, finansial, dan gaya hidup. Tahapan penelitian meliputi *preprocessing* data menggunakan *label encoding* dan normalisasi, dilanjutkan dengan proses *clustering* pada beberapa variasi nilai *weighting exponent*. Kualitas hasil *clustering* dievaluasi menggunakan *Xie-Beni Index* sebagai ukuran kekompakan dan keterpisahan *cluster*. Hasil menunjukkan bahwa nilai *weighting exponent* yang lebih kecil menghasilkan kualitas *cluster* yang lebih baik, sedangkan peningkatan nilai *weighting exponent* menyebabkan kenaikan nilai indeks. Kombinasi terbaik diperoleh pada tiga *cluster* dengan nilai *weighting exponent* sebesar 1,1 yang menghasilkan nilai *Xie-Beni Index* terendah. Hasil *clustering* mampu mengelompokkan data ke dalam tiga kategori risiko depresi, yaitu rendah, sedang, dan tinggi. Temuan ini menunjukkan bahwa pengaturan parameter *weighting exponent* pada algoritma *Fuzzy Gustafson-Kessel* berperan penting dalam meningkatkan kualitas *clustering*.

Kata kunci: *Fuzzy Gustafson-Kessel*; *Weighting Exponent*; *Xie-Beni Index*; *Clustering*; Depresi

1. Pendahuluan

Clustering merupakan teknik analisis data yang digunakan untuk mengelompokkan objek berdasarkan kesamaan karakteristiknya [1]. *Clustering* juga dikenal sebagai analisis *cluster* dalam statistik yang bertujuan membentuk kelompok berdasarkan karakteristik yang serupa [2]. Di era *big data*, teknik ini berperan penting dalam menyederhanakan data berskala besar menjadi kelompok yang lebih homogen sehingga pola data lebih mudah diidentifikasi. Salah satu metode *clustering* yang umum digunakan adalah *K-Means*. Metode ini menggunakan pendekatan *hard clustering*, di mana setiap objek hanya menjadi anggota satu *cluster* secara tegas [3]. Selain itu, *K-Means* sensitif terhadap inisialisasi *centroid* [4]. Oleh karena itu, dikembangkan pendekatan *fuzzy clustering* yang memungkinkan setiap objek memiliki derajat keanggotaan pada lebih dari satu *cluster* [5].

Fuzzy C-Means (FCM) merupakan metode *fuzzy clustering* yang populer [6]. Namun, FCM menggunakan jarak *Euclidean* yang mengasumsikan bentuk *cluster* sferis [6]. Pada data dengan distribusi tidak isotropik, performa FCM menjadi

kurang optimal [7]. Untuk mengatasi keterbatasan tersebut, digunakan algoritma *Fuzzy Gustafson-Kessel* (FGK) yang mampu mengakomodasi bentuk *cluster* yang lebih fleksibel [8]. Kinerja algoritma FGK sangat dipengaruhi oleh parameter *weighting exponent* yang mengatur tingkat kefuzzian derajat keanggotaan [6]. Pemilihan nilai *weighting exponent* yang tidak tepat dapat menghasilkan *cluster* yang terlalu tegas atau terlalu kabur [9]. Selain itu, hasil *clustering* juga dipengaruhi oleh banyaknya *cluster* yang umumnya ditentukan melalui evaluasi menggunakan indeks validitas seperti *Xie-Beni Index*.

Beberapa penelitian sebelumnya telah menerapkan algoritma *Fuzzy Gustafson-Kessel* (FGK) dalam berbagai kasus, seperti pengelompokan data pada bidang tertentu [10], pengelompokan wilayah berdasarkan indikator kesehatan [11], serta pengelompokan wilayah berdasarkan indikator kesejahteraan [12]. Hasil penelitian tersebut menunjukkan bahwa FGK mampu menangani data dengan karakteristik yang kompleks dan tidak berbentuk sferis. Namun, sebagian besar penelitian tersebut masih berfokus pada implementasi algoritma dan belum melakukan analisis maupun optimalisasi terhadap parameter *weighting exponent*, padahal parameter

tersebut berperan penting dalam mengatur tingkat keaburan derajat keanggotaan data pada proses *clustering*.

Di sisi lain, depresi merupakan gangguan mental yang dapat memengaruhi berbagai aspek kehidupan individu, khususnya pada pelajar yang rentan terhadap tekanan akademik dan sosial [13]. Risiko depresi pada pelajar dipengaruhi oleh berbagai faktor yang saling berkaitan, sehingga data yang dihasilkan cenderung bersifat multidimensional dan mengandung ketidakpastian [14]. Oleh karena itu, pendekatan *clustering*, khususnya berbasis *fuzzy* seperti FGK, menjadi relevan untuk digunakan dalam mengelompokkan tingkat risiko depresi pelajar.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis pengaruh variasi nilai *weighting exponent* pada algoritma *Fuzzy Gustafson-Kessel* terhadap kualitas hasil *clustering* data risiko depresi pelajar serta menentukan parameter optimal berdasarkan *Xie-Beni Index*. Kebaruan penelitian ini terletak pada analisis sistematis variasi nilai *weighting exponent* dalam FGK dengan mempertimbangkan banyaknya *cluster* sebagai parameter pendukung untuk memperoleh struktur *cluster* yang optimal.

2. Metode

Penelitian ini merupakan penelitian kuantitatif deskriptif yang mengolah data numerik berupa variabel risiko depresi pelajar serta hasil implementasi algoritma *Fuzzy Gustafson-Kessel* (FGK). Analisis dilakukan untuk menggambarkan kualitas hasil *clustering* dan menentukan parameter *weighting exponent* yang optimal menggunakan *Xie-Beni Index*.

Data yang digunakan merupakan data sekunder berupa *Depression Student Dataset* yang diperoleh dari platform *Kaggle*¹. *Dataset* ini terdiri dari 502 data dengan tujuh variabel yang mencakup aspek akademik, sosial, dan psikologis pelajar, yaitu umur (X_1), tekanan akademik (X_2), kepuasan belajar (X_3), jam belajar (X_4), stres finansial (X_5), durasi tidur (X_6), dan pola makan (X_7). *Dataset* ini juga telah melalui proses pembersihan data seperti penghapusan duplikasi dan *missing values*.

Dalam penelitian ini, proses analisis dilakukan melalui beberapa tahapan, yaitu analisis deskriptif, *preprocessing* data, implementasi algoritma FGK, serta evaluasi hasil *clustering*. Sebelum dilakukan proses *clustering*, data terlebih dahulu melalui tahap pra-pemrosesan untuk menyesuaikan skala dan format data sehingga hasil pengelompokan yang diperoleh menjadi lebih akurat.

2.1. Preprocessing data

Preprocessing data merupakan tahap awal yang dilakukan untuk menyiapkan data mentah agar siap dianalisis. Tahap ini diperlukan karena data sering memiliki perbedaan skala dan format, sehingga perlu disesuaikan agar proses analisis dapat menghasilkan hasil yang lebih akurat. Pada penelitian ini, *Preprocessing* data meliputi transformasi data kategorik dan normalisasi data.

2.1.1. Label Encoding

Label *encoding* merupakan proses transformasi data kategorik menjadi bentuk numerik dengan mengubah setiap kategori menjadi bilangan bulat [15]. Metode ini digunakan agar data kategori dapat diproses oleh algoritma *clustering*. Dalam penelitian ini, label *encoding* diterapkan pada variabel durasi tidur dan pola makan [16].

2.1.2. Normalisasi Data

Normalisasi data dilakukan untuk menyamakan skala nilai antar variabel numerik tanpa menghilangkan perbedaan karakteristik data [17]. Proses ini penting untuk mencegah dominasi variabel tertentu dalam perhitungan jarak pada algoritma *clustering*. Metode normalisasi yang digunakan adalah *Min-Max*, yaitu metode yang mengubah rentang nilai data ke dalam interval tertentu, umumnya $[0, 1]$. Normalisasi *Min-Max* dinyatakan sebagai:

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

dengan x' adalah nilai hasil normalisasi, x adalah nilai data asli, x_{\min} adalah nilai minimum, dan x_{\max} adalah nilai maksimum dari suatu variabel.

2.2. Algoritma Fuzzy Gustafson-Kessel

Algoritma *Fuzzy Gustafson-Kessel* (FGK) merupakan pengembangan dari metode *Fuzzy C-Means* yang mampu mengidentifikasi *cluster* dengan bentuk, ukuran, dan orientasi berbeda melalui penggunaan *fuzzy covariance matrix* pada setiap *cluster* [8]. Algoritma FGK dilakukan secara iteratif dengan memperbarui matriks keanggotaan, pusat *cluster*, dan matriks kovarian hingga mencapai kondisi konvergen. Adapun langkah-langkah algoritma FGK sebagai berikut [6][18]:

2.2.1. Inisialisasi Matriks Keanggotaan

Matriks keanggotaan awal (U_0) dibangkitkan secara acak, di mana setiap elemen $\mu_{i,k}$ menyatakan derajat keanggotaan data ke- i terhadap *cluster* ke- k , dengan $i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$. Matriks keanggotaan tersebut dapat dituliskan sebagai:

$$U_0 = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,c} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n,1} & \mu_{n,2} & \cdots & \mu_{n,c} \end{bmatrix} \quad (2)$$

dengan:

$$\sum_{k=1}^c \mu_{i,k} = 1 \quad (3)$$

Pada persamaan tersebut, $\mu_{i,k}$ menyatakan derajat keanggotaan data ke- i terhadap *cluster* ke- k , dengan i sebagai indeks data dan k sebagai indeks *cluster*.

2.2.2. Pusat Cluster

Pusat *cluster* ditentukan berdasarkan derajat keanggotaan setiap data terhadap masing-masing *cluster*. Nilai pusat *cluster*

¹<https://www.kaggle.com/datasets/ikynahidwin/depression-student-dataset>

ke- k pada variabel ke- j dihitung menggunakan persamaan:

$$v_{k,j} = \frac{\sum_{i=1}^n (\mu_{i,k})^m x_{i,j}}{\sum_{i=1}^n (\mu_{i,k})^m} \quad (4)$$

Pusat *cluster* untuk setiap *cluster* kemudian dinyatakan dalam bentuk vektor:

$$v_k = (v_{k,1}, v_{k,2}, \dots, v_{k,p}) \quad (5)$$

Pada persamaan tersebut, $v_{k,j}$ menyatakan nilai pusat *cluster* ke- k pada variabel ke- j , $\mu_{i,k}$ menyatakan derajat keanggotaan data ke- i terhadap *cluster* ke- k , $x_{i,j}$ menyatakan nilai data ke- i pada variabel ke- j , dan m menyatakan *weighting exponent*.

2.2.3. Matriks Kovarian

Matriks kovarian pada setiap *cluster* digunakan untuk menyesuaikan bentuk, ukuran, dan orientasi *cluster* terhadap struktur data. Matriks kovarian untuk *cluster* ke- k dihitung menggunakan persamaan:

$$F_k = \left(\sum_{i=1}^n (\mu_{i,k})^m (x_i - v_k)(x_i - v_k)^T \right) \left(\sum_{i=1}^n (\mu_{i,k})^m \right)^{-1} \quad (6)$$

Matriks kovarian ini kemudian digunakan dalam pembentukan *norm inducing matrix* untuk menghitung jarak antar data dan pusat *cluster*. Pada persamaan tersebut, F_k menyatakan matriks kovarian *cluster* ke- k , x_i menyatakan vektor data ke- i , v_k menyatakan pusat *cluster* ke- k , $\mu_{i,k}$ menyatakan derajat keanggotaan data ke- i terhadap *cluster* ke- k , dan m menyatakan *weighting exponent*.

2.2.4. Jarak Mahalanobis

Jarak antara data dan pusat *cluster* dihitung menggunakan jarak *Mahalanobis* yang mempertimbangkan struktur kovarian setiap *cluster*. Jarak tersebut dinyatakan sebagai:

$$D_{i,kA_k}^2 = (x_i - v_k)^T A_k (x_i - v_k) \quad (7)$$

$$A_k = \left[\rho_k \det(F_k)^{\frac{1}{p}} F_k^{-1} \right] \quad (8)$$

Pada Equation 7, simbol T menyatakan operasi transpose vektor. Operasi ini diperlukan agar dimensi matriks pada perhitungan jarak *Mahalanobis* menjadi sesuai, yaitu dengan mengubah vektor kolom menjadi vektor baris sehingga hasil perkalian matriks menghasilkan nilai skalar. Sementara itu, notasi $\det(F_k)^{1/p}$ pada Equation 8 menyatakan operasi perpangkatan terhadap determinan matriks kovarian F_k . Nilai determinan tersebut dipangkatkan sebesar $1/p$ untuk melakukan normalisasi terhadap bentuk dan volume *cluster*.

2.2.5. Fungsi Objektif

Algoritma *Fuzzy Gustafson-Kessel* meminimalkan fungsi objektif pada setiap iterasi yang dinyatakan sebagai:

$$J_{GK}^{(t)} = \sum_{i=1}^n \sum_{k=1}^c (\mu_{i,k})^m D_{i,kA_k}^2 \quad (9)$$

Pada persamaan tersebut, $J_{GK}^{(t)}$ menyatakan fungsi objektif pada iterasi ke- t , $\mu_{i,k}$ menyatakan derajat keanggotaan data ke- i terhadap *cluster* ke- k , m menyatakan *weighting exponent*, dan D_{i,kA_k}^2 menyatakan jarak antara data ke- i dan pusat *cluster* ke- k berdasarkan matriks A_k .

2.2.6. Pembaruan Matriks Keanggotaan

Matriks keanggotaan diperbarui pada setiap iterasi berdasarkan jarak antara data dan pusat *cluster*. Pembaruan derajat keanggotaan dinyatakan sebagai:

$$\mu_{i,k}^{(t+1)} = \left[\sum_{q=1}^c \left(\frac{D_{i,kA_k}}{D_{i,qA_q}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (10)$$

Pada persamaan tersebut, $\mu_{i,k}$ menyatakan derajat keanggotaan data ke- i terhadap *cluster* ke- k , D_{i,kA_k} dan D_{i,qA_q} menyatakan jarak data ke- i terhadap *cluster* ke- k dan ke- q , serta m menyatakan *weighting exponent*.

2.2.7. Kriteria Konvergensi

Proses iterasi dilakukan hingga memenuhi kriteria konvergensi, yaitu ketika nilai norma Frobenius dari selisih matriks keanggotaan pada dua iterasi berturut-turut lebih kecil dari nilai ambang batas yang telah ditentukan. Kondisi tersebut dinyatakan sebagai:

$$\|U_{t+1} - U_t\| < \varepsilon \quad (11)$$

Pada persamaan tersebut, U_t menyatakan matriks keanggotaan pada iterasi ke- t , sedangkan U_{t+1} menyatakan matriks keanggotaan pada iterasi berikutnya. Simbol $\|\cdot\|$ menyatakan norma Frobenius yang digunakan untuk mengukur besarnya perubahan antara dua matriks keanggotaan. Proses iterasi dinyatakan konvergen apabila nilai norma dari selisih kedua matriks tersebut lebih kecil dari nilai ambang batas ε , dengan ε merupakan nilai kecil yang mendekati nol.

2.3. Xie-Beni Index

Evaluasi kualitas hasil *clustering* dilakukan menggunakan *Xie-Beni Index* (XB), yaitu ukuran validitas *cluster* yang membandingkan tingkat kekompakan internal *cluster* dan jarak antar pusat *cluster* [19]. Semakin kecil nilai XB, maka *cluster* yang dihasilkan semakin kompak dan memiliki pemisahan yang lebih baik. *Xie-Beni Index* dinyatakan sebagai:

$$v_{XB} = \frac{S}{d_{\min}} \quad (12)$$

Pada persamaan tersebut, v_{XB} menyatakan nilai *Xie-Beni Index*, S menyatakan tingkat kekompakan *cluster*, dan d_{\min} menyatakan jarak minimum antar pusat *cluster*.

2.4. Principal Component Analysis

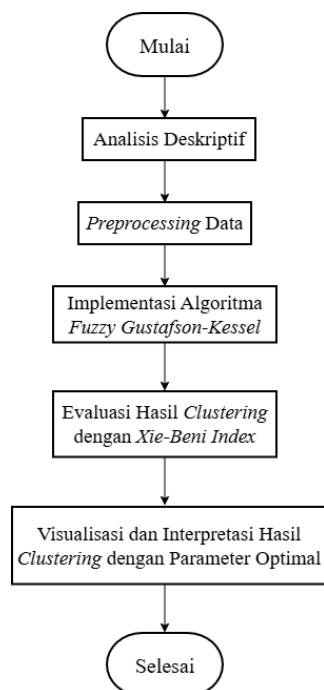
Principal Component Analysis (PCA) merupakan metode reduksi dimensi yang mentransformasikan data ke dalam beberapa komponen utama yang saling ortogonal sehingga dapat merepresentasikan sebagian besar variasi data [20]. Proses PCA dilakukan melalui normalisasi data, pembentukan matriks kovarians, serta perhitungan nilai eigen dan vektor eigen dengan persamaan:

$$C \cdot v = \lambda \cdot v \quad (13)$$

di mana C menyatakan matriks kovarians, v menyatakan vektor eigen, dan λ menyatakan nilai eigen [21]. Dalam penelitian ini, PCA digunakan untuk memvisualisasikan hasil *clustering* sehingga pola dan keterpisahan antar *cluster* dapat diamati dengan lebih jelas.

2.5. Diagram Alir Penelitian

Alur penelitian ditunjukkan pada Gambar 1. Tahapan penelitian meliputi pengumpulan data, *preprocessing* (label *encoding* dan normalisasi), implementasi algoritma *Fuzzy Gustafson-Kessel* (FGK), evaluasi menggunakan *Xie-Beni Index* untuk menentukan parameter optimal, serta visualisasi dan interpretasi hasil *clustering* menggunakan PCA.



Gambar 1: Diagram Alir Penelitian

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil analisis variasi nilai *weighting exponent* (m) terhadap kualitas hasil *clustering* menggunakan algoritma *Fuzzy Gustafson-Kessel* (FGK) pada data risiko depresi pelajar. Proses analisis meliputi tahap *preprocessing* data, implementasi algoritma, evaluasi menggunakan *Xie-Beni Index*, serta visualisasi hasil *clustering*.

3.1. Analisis Deskriptif

Analisis deskriptif dilakukan untuk memberikan gambaran umum karakteristik data risiko depresi pelajar yang digunakan dalam penelitian ini. Data terdiri dari 502 responden dengan tujuh variabel, yaitu lima variabel numerik dan dua variabel kategorikal yang ditunjukkan pada Tabel 1 dan Tabel 2.

Tabel 1: Statistik Deskriptif Variabel Numerik

Variabel	Mean	Median	Standar Deviasi	Minimum	Maximum
X_1	26,24	26,5	4,90	18	34
X_2	3,00	3,00	1,39	1	5
X_3	3,08	3,00	1,37	1	5
X_4	6,40	7,00	3,74	0	12
X_5	2,93	3,00	1,43	1	5

Berdasarkan Tabel 1, variabel umur memiliki rata-rata sebesar 26,24 tahun. Variabel tekanan akademik berada pada nilai tengah skala dengan rata-rata 3,00, sedangkan

jam belajar memiliki rata-rata 6,40 jam yang menunjukkan adanya variasi antar responden.

Tabel 2: Analisis Deskriptif Variabel Kategorikal

Variabel	Kategori	Frekuensi	Persentase
X_6	Kurang dari 5 jam	123	25%
	5–6 jam	123	25%
	7–8 jam	128	25%
	Lebih dari 8 jam	128	25%
Total		502	100%
X_7	Sehat	161	32%
	Sedang	172	34%
	Tidak Sehat	169	34%
Total		502	100%

Berdasarkan Tabel 2, variabel durasi tidur memiliki distribusi yang relatif merata pada setiap kategori. Pada variabel pola makan, kategori sedang dan tidak sehat mendominasi dengan proporsi sekitar 34%, menunjukkan bahwa sebagian besar responden belum memiliki pola makan yang optimal.

3.2. Preprocessing Data

3.2.1. Label Encoding

Label *encoding* diterapkan pada variabel kategorikal, yaitu durasi tidur dan pola makan, untuk mengubah data menjadi bentuk numerik sehingga dapat diproses oleh algoritma. Skema pengkodean yang digunakan ditunjukkan pada Tabel 3.

Tabel 3: Skema Label Encoding

X_1	X_2	X_3	X_4	X_5	X_6	X_7
28	2	4	9	2	2	1
28	4	5	7	1	1	2
25	1	3	10	4	1	0
23	1	4	7	2	3	0
31	1	5	4	2	3	2

Berdasarkan Tabel 3, setiap kategori diubah menjadi bilangan bulat secara berurutan, sehingga seluruh variabel dalam dataset telah berada dalam format numerik.

3.2.2. Normalisasi Data

Normalisasi dilakukan untuk menyamakan skala antar variabel agar tidak terjadi dominasi dalam perhitungan jarak. Metode yang digunakan adalah *Min-Max Normalization* yang mengubah nilai data ke dalam rentang 0 hingga 1. Adapun contoh hasil untuk lima data pertama ditunjukkan pada Tabel 4.

Tabel 4: Contoh Dataset setelah Normalisasi Min-Max

X_1	X_2	X_3	X_4	X_5	X_6	X_7
0,625	0,25	0,75	0,75	0,25	0,6667	0,5
0,625	0,75	1	0,5833	0	0,3333	1
0,4375	0	0,5	0,8333	0,75	0,3333	0
0,3125	0	0,75	0,5833	0,25	1	0
0,8125	0	1	0,3333	0,25	1	1

Berdasarkan Tabel 4, seluruh variabel telah berada pada rentang 0 hingga 1, sehingga data dapat digunakan dalam proses *clustering* menggunakan algoritma *Fuzzy Gustafson-Kessel*.

3.3. Implementasi Algoritma Fuzzy Gustafson-Kessel

Data yang digunakan merupakan data risiko depresi pelajar yang telah melalui tahap *preprocessing*, yaitu label *encoding* dan normalisasi *Min-Max*. Data tersebut direpresentasikan dalam bentuk matriks X berukuran $n \times p$, dengan n menyatakan jumlah data dan p menyatakan jumlah variabel. Setiap baris merepresentasikan satu responden, sedangkan setiap kolom menunjukkan nilai variabel hasil normalisasi.

$$X = \begin{bmatrix} 0,625 & 0,25 & 0,75 & 0,75 & \dots & 0,6667 & 0,5 \\ 0,625 & 0,75 & 1 & 0,5833 & \dots & 0,3333 & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0,9375 & 0,75 & 0,75 & 0,6667 & \dots & 1 & 1 \\ 0 & 1 & 0,5 & 0,5 & \dots & 1 & 0 \end{bmatrix}$$

Sebelum proses iterasi dilakukan, parameter algoritma *Fuzzy Gustafson-Kessel* ditentukan terlebih dahulu, yaitu jumlah *cluster* ($c = 2, 3, 4, 5$), *weighting exponent* ($m = 1,1; 1,3; 1,5; 1,7; 1,9; 2,1$), maksimum iterasi sebesar 1000, ambang konvergensi ($\epsilon = 0,0001$), serta nilai awal fungsi objektif sebesar 0.

Weighting exponent atau pangkat pembobot (m) digunakan untuk mengatur tingkat kefuzzian dalam proses *clustering*, sehingga variasi nilai m dapat memengaruhi hasil *cluster* yang terbentuk. Pada penelitian ini, nilai m divariasikan pada rentang $m > 1$, dengan nilai awal $m = 1,1$ dipilih karena merupakan nilai yang mendekati 1 dan masih memenuhi syarat perhitungan algoritma. Selanjutnya, variasi nilai m disusun secara bertahap dengan interval yang konsisten untuk memudahkan analisis pengaruh perubahan *weighting exponent* terhadap kualitas hasil *clustering*.

Tahapan implementasi algoritma dilakukan berdasarkan parameter yang telah ditetapkan. Pada bagian ini, penyajian proses perhitungan difokuskan pada kombinasi parameter $c = 2$ dan $m = 1,1$ sebagai representasi proses iteratif algoritma, sedangkan perhitungan dilakukan untuk seluruh variasi parameter.

3.3.1. Inisialisasi Matriks Keanggotaan

Matriks keanggotaan awal (U_0) dibentuk melalui pembangkitan bilangan acak menggunakan program *Python*, dengan $i = 1, 2, \dots, 502$ dan $k = 1, 2$. Setiap entri $\mu_{i,k}$ merepresentasikan derajat keanggotaan data ke- i terhadap *cluster* ke- k , dengan syarat bahwa jumlah nilai pada setiap baris bernilai 1. Hasil pembangkitan matriks keanggotaan awal untuk $c = 2$ ditunjukkan sebagai berikut:

$$U_0 = \begin{bmatrix} 0,6831 & 0,3619 \\ 0,5518 & 0,4482 \\ \vdots & \vdots \\ 0,1193 & 0,8807 \\ 0,4586 & 0,5414 \end{bmatrix}$$

3.3.2. Pusat Cluster

Pusat *cluster* dihitung menggunakan Equation 4 berdasarkan matriks keanggotaan dan nilai $m = 1,1$. Hasil pusat *cluster* untuk $c = 2$ disajikan pada Tabel 5.

Tabel 5: Hasil Pusat Dua *Cluster* ($m = 1,1$)

Variabel	Cluster 1	Cluster 2
X_1	0,5180	0,5127
X_2	0,4916	0,5095
X_3	0,5063	0,5297
X_4	0,5393	0,5286
X_5	0,4829	0,4814
X_6	0,4990	0,5133
X_7	0,5049	0,4791

Pusat *cluster* tersebut dapat dinyatakan dalam bentuk vektor sebagai berikut:

$$v_1 = (0,5180, 0,4916, 0,5063, 0,5393, 0,4829, 0,4990, 0,5049)$$

$$v_2 = (0,5127, 0,5095, 0,5297, 0,5286, 0,4814, 0,5133, 0,4791)$$

3.3.3. Matriks Kovarian

Matriks kovarian dihitung menggunakan Equation 6 untuk merepresentasikan bentuk dan penyebaran data pada setiap *cluster*. Karena jumlah variabel yang digunakan sebanyak 7, maka matriks kovarian yang dihasilkan berukuran 7×7 . Hasil perhitungan matriks kovarian untuk $c = 2$ dan $m = 1,1$ adalah sebagai berikut:

$$F_1 = \begin{bmatrix} 0,0922 & 0,0033 & -0,0042 & 0,0063 & -0,0111 & -0,0057 & 0,0085 \\ 0,0033 & 0,1236 & -0,0137 & 0,0100 & 0,0025 & -0,0049 & 0,0065 \\ -0,0042 & -0,0137 & 0,1183 & -0,0039 & -0,0011 & 0,0001 & -0,0030 \\ 0,0063 & 0,0100 & -0,0039 & 0,1017 & 0,0039 & -0,0103 & 0,0050 \\ -0,0111 & 0,0025 & -0,0011 & 0,0039 & 0,1254 & -0,0043 & -0,0028 \\ -0,0057 & -0,0049 & 0,0001 & -0,0103 & -0,0043 & 0,1457 & -0,0230 \\ 0,0085 & 0,0065 & -0,0030 & 0,0050 & -0,0028 & -0,0230 & 0,1637 \end{bmatrix}$$

$$F_2 = \begin{bmatrix} 0,0946 & 0,0025 & 0,0026 & 0,0032 & -0,0095 & -0,0047 & 0,0069 \\ 0,0025 & 0,1176 & -0,0107 & 0,0021 & -0,0016 & 0,0068 & -0,0017 \\ 0,0026 & -0,0107 & 0,1168 & 0,0006 & -0,0080 & -0,0047 & 0,0090 \\ 0,0032 & 0,0021 & 0,0006 & 0,0926 & -0,0004 & -0,0055 & 0,0048 \\ -0,0095 & -0,0016 & -0,0080 & -0,0004 & 0,1277 & -0,0045 & -0,0069 \\ -0,0047 & 0,0068 & -0,0047 & -0,0055 & -0,0045 & 0,1325 & -0,0155 \\ 0,0069 & -0,0017 & 0,0090 & 0,0048 & -0,0069 & -0,0155 & 0,1647 \end{bmatrix}$$

3.3.4. Jarak Mahalanobis

Jarak *Mahalanobis* dihitung menggunakan Equation 7 dengan memanfaatkan *norm-inducing matrix* A_k yang diperoleh dari matriks kovarian melalui Equation 8. Karena jumlah variabel sebanyak 7, maka matriks A_k yang dihasilkan berukuran 7×7 . Hasil perhitungan matriks A_k untuk $c = 2$ dan $m = 1,1$ adalah sebagai berikut:

$$A_1 = \begin{bmatrix} 1,3423 & -0,0226 & 0,0415 & -0,0777 & 0,1221 & 0,0412 & -0,0574 \\ -0,0226 & 1,0013 & 0,1113 & -0,0883 & -0,0182 & 0,0210 & -0,0311 \\ 0,0415 & 0,1113 & 1,0379 & 0,0260 & 0,0100 & 0,0092 & 0,0134 \\ -0,0777 & -0,0883 & 0,0260 & 1,2137 & -0,0407 & 0,0752 & -0,0195 \\ 0,1221 & -0,0182 & 0,0100 & -0,0407 & 0,9786 & 0,0331 & 0,0172 \\ 0,0412 & 0,0210 & 0,0092 & 0,0752 & 0,0331 & 0,8570 & 0,1158 \\ -0,0574 & -0,0310 & 0,0134 & -0,0195 & 0,0172 & 0,1158 & 0,7602 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1,2643 & -0,0303 & -0,0194 & -0,0380 & 0,0911 & 0,0419 & -0,0436 \\ -0,0303 & 1,0162 & 0,0930 & -0,0255 & 0,0144 & -0,0499 & 0,0034 \\ -0,0194 & 0,0930 & 1,0281 & -0,0034 & 0,0626 & 0,0271 & -0,0490 \\ -0,0380 & -0,0255 & -0,0034 & 1,2808 & 0,0005 & 0,0498 & -0,0309 \\ 0,0911 & 0,0144 & 0,0626 & 0,0005 & 0,9378 & 0,0406 & 0,0362 \\ 0,0419 & -0,0499 & 0,0271 & 0,0498 & 0,0406 & 0,9084 & 0,0821 \\ -0,0436 & 0,0034 & -0,0490 & -0,0309 & 0,0362 & 0,0821 & 0,7310 \end{bmatrix}$$

Berdasarkan matriks tersebut, jarak *Mahalanobis* antara data dan pusat *cluster* dihitung. Hasil jarak untuk $c = 2$ disajikan pada Tabel 6.

Tabel 6: Hasil Jarak Mahalanobis Dua Cluster ($m = 1,1$)

Data	Cluster 1	Cluster 2
1	0,5126	0,5073
2	0,8708	0,8161
3	0,8116	0,8069
⋮	⋮	⋮
502	1,0184	0,9648

3.3.5. Fungsi Objektif

Nilai fungsi objektif dihitung menggunakan Equation 9 sebagai indikator kualitas hasil clustering pada setiap iterasi. Untuk kasus ilustrasi $c = 2$ dan $m = 1,1$, diperoleh:

$$\text{Cluster 1} = 192,3949$$

$$\text{Cluster 2} = 204,3948$$

Sehingga nilai fungsi objektif pada iterasi ke-1 adalah:

$$J_{GK}^{(1)} = 396,7897$$

Nilai ini akan diperbarui pada iterasi berikutnya hingga memenuhi kriteria konvergensi.

3.3.6. Pembaruan Matriks Keanggotaan

Matriks keanggotaan diperbarui menggunakan Equation 10 untuk memperoleh nilai derajat keanggotaan pada iterasi berikutnya. Untuk ilustrasi $c = 2$ dan $m = 1,1$, contoh hasil pembaruan pada data pertama adalah:

$$\text{Cluster 1} = 0,4485$$

$$\text{Cluster 2} = 0,5515$$

Hasil matriks keanggotaan baru ditampilkan pada Tabel 7.

Tabel 7: Hasil Matriks Keanggotaan Baru dengan Dua Cluster ($m = 1,1$)

Data	Cluster 1	Cluster 2
1	0,4485	0,5515
2	0,2144	0,7856
3	0,4713	0,5287
⋮	⋮	⋮
502	0,2531	0,7469

3.3.7. Kriteria Konvergensi

Konvergensi ditentukan dengan membandingkan perubahan matriks keanggotaan antar iterasi, yaitu U_t dan U_{t+1} , menggunakan norma selisih matriks. Untuk ilustrasi $c = 2$ dan $m = 1,1$, diperoleh:

$$\|U_{t+1} - U_t\| = 7,9552$$

Nilai tersebut lebih besar dari ambang batas $\varepsilon = 0,0001$, sehingga algoritma belum konvergen dan iterasi dilanjutkan. Setelah proses iterasi berlangsung, pada iterasi ke-54 diperoleh:

$$\|U_{t+1} - U_t\| = 0,000091523$$

Karena nilainya sudah lebih kecil dari ε , maka algoritma dinyatakan konvergen dan proses clustering dihentikan.

3.3.8. Hasil Clustering

Setelah algoritma Fuzzy Gustafson-Kessel dengan $m = 1,1$ mencapai kondisi konvergen, penentuan cluster dilakukan berdasarkan nilai derajat keanggotaan terbesar pada setiap data. Data akan masuk ke cluster dengan nilai keanggotaan dominan. Hasil clustering untuk $c = 2$ disajikan pada Tabel 8.

Tabel 8: Hasil Clustering Dua Cluster ($m = 1,1$)

Data	Derajat Keanggotaan		Cluster
	1	2	
1	0,00000006	0,99999994	2
2	0,00000003	0,99999997	2
3	0,8820	0,1180	1
⋮	⋮	⋮	⋮
502	0,9592	0,0408	1

3.4. Evaluasi Hasil Clustering

Evaluasi hasil clustering dilakukan menggunakan Xie-Beni Index (XB) untuk membandingkan kualitas cluster pada setiap kombinasi parameter weighting exponent (m) dan banyaknya cluster (c). Nilai XB yang lebih kecil menunjukkan cluster yang lebih kompak dan memiliki pemisahan yang lebih baik. Perhitungan nilai Xie-Beni Index dilakukan menggunakan Equation 12. Hasil perhitungan untuk seluruh kombinasi parameter disajikan pada Tabel 9.

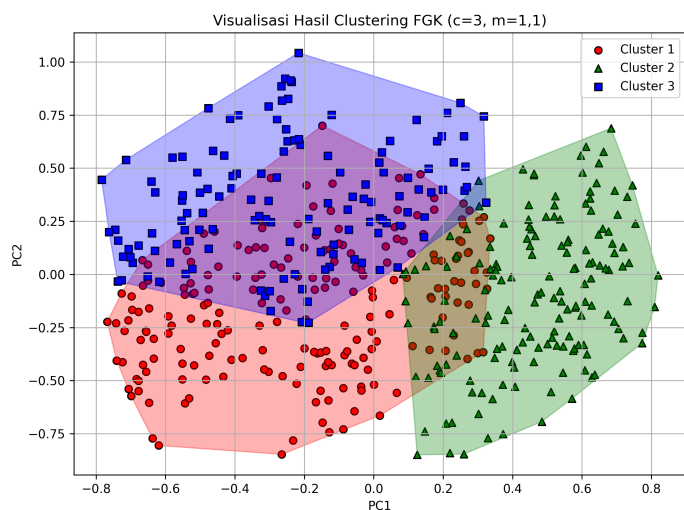
Tabel 9: Hasil Evaluasi Clustering dengan Xie-Beni Index

Banyaknya Cluster (c)	Weighting Exponent (m)	Nilai Xie-Beni Index
2	1,1	2,1243402611
	1,3	2,0795837205
	1,5	13,5108918119
	1,7	262407710,05642
	1,9	767503590,26202
	2,1	1103379379,3075
3	1,1	1,5542301750
	1,3	1,7156973536
	1,5	8,6214837453
	1,7	713761178,63209
	1,9	1496369014,1610
	2,1	1899745079,1890
4	1,1	2,1092060904
	1,3	1,8476430912
	1,5	6,8329067408
	1,7	3071516307,3882
	1,9	3256546739,4304
	2,1	2190180230,1422
5	1,1	2,9550053889
	1,3	6,5562159904
	1,5	6,2393741202
	1,7	1457974274,5305
	1,9	2576713854,7351
	2,1	2693981637,9315

Berdasarkan Tabel 9, terlihat bahwa nilai Xie-Beni Index (XB) cenderung meningkat seiring bertambahnya nilai weighting exponent (m), terutama pada $m \geq 1,5$ yang menunjukkan penurunan kualitas cluster. Nilai XB terkecil diperoleh pada kombinasi $c = 3$ dan $m = 1,1$ sebesar 1,5542. Selain itu, nilai XB pada $m = 1,3$ juga relatif rendah pada beberapa variasi cluster. Dengan demikian, dapat disimpulkan bahwa kombinasi parameter optimal dalam penelitian ini adalah $c = 3$ dan $m = 1,1$.

3.5. Visualisasi dan Interpretasi Hasil Clustering dengan Parameter Optimal

Hasil *clustering* dengan parameter optimal ($c = 3$ dan $m = 1,1$) selanjutnya divisualisasikan menggunakan *Principal Component Analysis* (PCA) untuk melihat pola sebaran data dan pemisahan antar *cluster* pada ruang dua dimensi. Visualisasi ini merupakan hasil proyeksi data berdimensi tujuh ke dalam dua komponen utama yang mampu merepresentasikan sebagian besar variasi data. Adapun visualisasi dengan parameter optimal ditunjukkan pada Gambar 2.



Gambar 2: Visualisasi Hasil Clustering

Berdasarkan Gambar 2, terlihat bahwa data terbagi ke dalam tiga *cluster* dengan pemisahan yang cukup jelas. Jumlah anggota pada masing-masing *cluster* adalah 209 data pada *cluster* 1, 161 data pada *cluster* 2, dan 132 data pada *cluster* 3. Untuk menginterpretasikan karakteristik tiap *cluster*, dilakukan analisis terhadap nilai rata-rata masing-masing variabel, yang disajikan pada Tabel 10.

Tabel 10: Nilai Rata-rata Variabel Tiap Cluster

Cluster	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	0,5374	0,4904	0,5036	0,5179	0,2380	0,5502	0,2608
2	0,5198	0,5078	0,5264	0,5487	0,4674	0,4534	1
3	0,4740	0,5095	0,5341	0,5404	0,8864	0,5025	0,2386

Berdasarkan Tabel 10, setiap *cluster* memiliki karakteristik yang berbeda. *Cluster* 1 memiliki nilai rata-rata variabel yang cenderung berada pada tingkat sedang, dengan stres finansial relatif rendah (0,2380), sehingga dikategorikan sebagai risiko depresi sedang. *Cluster* 2 menunjukkan kondisi yang lebih baik dengan nilai kepuasan belajar (0,5264), jam belajar (0,5487), dan pola makan tertinggi (1), sehingga dikategorikan sebagai risiko depresi rendah. *Cluster* 3 memiliki nilai stres finansial paling tinggi (0,8864) disertai tekanan akademik yang cukup tinggi, sehingga dikategorikan sebagai risiko depresi tinggi.

4. Kesimpulan

Penelitian ini menganalisis pengaruh variasi nilai *weighting exponent* (m) pada algoritma *Fuzzy Gustafson-Kessel* (FGK) terhadap kualitas *clustering* data risiko depresi pelajar. Hasil

penelitian menunjukkan bahwa nilai m berpengaruh terhadap kualitas *cluster*, di mana nilai m yang lebih kecil menghasilkan *cluster* yang lebih baik, sedangkan nilai m yang lebih besar menurunkan kualitas akibat meningkatnya tingkat kekaburan. Parameter optimal diperoleh pada kombinasi $c = 3$ dan $m = 1,1$ berdasarkan nilai *Xie-Beni Index* minimum.

Dengan parameter tersebut, data pelajar berhasil dikelompokkan ke dalam tiga kategori risiko depresi, yaitu rendah, sedang, dan tinggi, sehingga menunjukkan bahwa optimalisasi parameter FGK mampu menghasilkan pengelompokan yang lebih representatif. Penelitian selanjutnya disarankan untuk mengkaji variasi parameter lain seperti *threshold* (ε) serta menggunakan indeks validitas tambahan untuk memperoleh hasil yang lebih komprehensif.

Pernyataan Kontribusi Penulis (CRediT)

Amira Adelia Putri: Konseptualisasi, Metodologi, Penulisan Draf Awal. **Intan Nisfulaila:** Kurasi Data, Analisis Formal, Penulisan Telaah dan Penyuntingan. **Erna Herawati:** Supervisi, Validasi.

Deklarasi Penggunaan AI atau Teknologi Berbasis AI

Model ChatGPT-5.5 digunakan untuk membantu penyusunan draf awal, perbaikan struktur kalimat, dan penyuntingan bahasa pada penulisan naskah ini.

Deklarasi Konflik Kepentingan

Penulis menyatakan tidak ada konflik kepentingan.

Pendanaan dan Ucapan Terima Kasih

Penelitian ini tidak menerima pendanaan eksternal. Penulis menyampaikan terima kasih kepada semua pihak yang telah memberikan saran dan dukungan selama proses penelitian ini.

Ketersediaan Data

Student Depression Dataset yang digunakan dalam penelitian ini tersedia secara publik melalui repositori Kaggle dan telah dicantumkan pada Section 2.

Daftar Pustaka

- [1] P. W. Rahayu et al., *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024. Available online.
- [2] M. W. Talakua, Z. A. Leleury, and A. W. Talluta, "Analisis cluster dengan menggunakan metode k-means untuk pengelompokan kabupaten/kota di provinsi maluku berdasarkan indeks pembangunan tahun 2024," *Barekeng Jurnal Ilmu Matematika dan Terapan*, vol. 11, no. 2, pp. 119–128, 2017. DOI: [10.30598/barekengvol11iss2pp119-128](https://doi.org/10.30598/barekengvol11iss2pp119-128).

- [3] F. H. Wagner et al., “K-textures, a self-supervised hard clustering deep learning algorithm for satellite image segmentation,” *Frontiers in Environmental Science*, vol. 10, 2022. DOI: [10.3389/fenvs.2022.946729](https://doi.org/10.3389/fenvs.2022.946729).
- [4] Y. Duan, Q. Liu, and S. Xia, “An improved initialization center k-means clustering algorithm based on distance and density,” *AIP Conference Proceedings*, vol. 1955, no. 1, 2018. DOI: [10.1063/1.5033710](https://doi.org/10.1063/1.5033710).
- [5] L. Zadeh, “Fuzzy sets*,” *Journal of Information and Control*, vol. 8, no. 3, pp. 338–353, 1965. DOI: [10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- [6] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, 1981. DOI: [10.1007/9781475704501](https://doi.org/10.1007/9781475704501).
- [7] Z. Cebeci and F. Yildiz, “Comparison of k-means and fuzzy c-means algorithms on different cluster structures,” *Journal of Agricultural Informatics*, vol. 6, no. 3, pp. 13–23, 2015. DOI: [10.17700/jai.2015.6.3.196](https://doi.org/10.17700/jai.2015.6.3.196).
- [8] D. E. Gustafson and W. C. Kessel, “Fuzzy clustering with a fuzzy covariance matrix,” in *Proceedings of the IEEE Conference on Decision and Control*, 1978. DOI: [10.1109/CDC.1978.268028](https://doi.org/10.1109/CDC.1978.268028).
- [9] N. R. Pal and J. C. Bezdek, “On cluster validity for the fuzzy c-means model,” *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370–379, 1995. DOI: [10.1109/91.413225](https://doi.org/10.1109/91.413225).
- [10] F. A. A. Putri and N. Rochmawati, “Penerapan algoritma gustafson-kessel untuk clustering tweets mention akun go-jek dan grab indonesia,” *Journal of Informatics and Computer Science (JINACS)*, vol. 2, no. 3, pp. 204–212, 2021. DOI: [10.26740/jinacs.v2n03.p204-212](https://doi.org/10.26740/jinacs.v2n03.p204-212).
- [11] M. Fajri, R. Rais, and L. Handayani, “Regions Grouping in Central Sulawesi Province by Transmitted Disease Using Fuzzy Gustafson Kessel,” *BAREKENG: Journal of Mathematics and Its Applications*, vol. 17, no. 1, pp. 1–10, 2023. Available online.
- [12] T. Hidayat, “Analisis cluster dengan metode ensemble rock dalam pengelompokan kabupaten/kota di provinsi sulawesi selatan berdasarkan indikator kesejahteraan rakyat,” M.S. thesis, Universitas Negeri Makassar, Makassar, 2022.
- [13] World Health Organization, *Depressive disorder (depression)*, Diakses pada 7 September 2025, 2025. Available online.
- [14] H. Han, “Fuzzy clustering algorithm for university students’ psychological fitness and performance detection,” *Heliyon*, vol. 12, no. 5, 2023. DOI: [10.1016/j.heliyon.2023.e18550](https://doi.org/10.1016/j.heliyon.2023.e18550).
- [15] J. K. Wororomi et al., *Data Mining (Memahami Pola di Balik Angka)*. Eureka Media Aksara, 2024. Available online.
- [16] C. Herdian, A. Kamila, and I. G. A. M. Budidarma, “Studi kasus feature engineering untuk data teks: Perbandingan label encoding dan one-hot encoding pada metode linear regresi,” *Technologia : Jurnal Ilmiah*, vol. 15, no. 1, pp. 93–108, 2024. DOI: [10.31602/tji.v15i1.13457](https://doi.org/10.31602/tji.v15i1.13457).
- [17] R. K. Hapsari et al., *Data Mining*. Indie Press, 2024. Available online.
- [18] Y. S. Paradilla, M. N. Hayati, and Sifriyani, “Implementation of the fuzzy gustafson-kessel method on grouping districts/cities in kalimantan island based on poverty issues factors. barekeng,” *BAREKENG: Journal of Mathematics and Its Applications*, vol. 17, no. 1, pp. 0125–0134, 2023. DOI: [10.30598/barekengvol17iss1pp0125-0134](https://doi.org/10.30598/barekengvol17iss1pp0125-0134).
- [19] Z. Feng and J. Fan, “A novel validity index in fuzzy clustering algorithm,” *International Journal of Wireless and Mobile Computing*, vol. 10, no. 2, p. 183, 2016. DOI: [10.1504/IJWMC.2016.076153](https://doi.org/10.1504/IJWMC.2016.076153).
- [20] A. R. Firdawanti, H. F. Ahmad, and N. Agustiani, “Analisis perbandingan kinerja algoritma k-means dan k-medoids dengan reduksi dimensi pca pada indikator kesehatan dan sosial,” *Bulletin of Computer Science Research*, vol. 5, no. 5, pp. 1235–1247, 2025. DOI: [10.47065/bulletincsr.v5i5.742](https://doi.org/10.47065/bulletincsr.v5i5.742).
- [21] Syahreini et al., “Analisis pola data klimatologi potensial dengan principal component analysis,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 6, 2025. DOI: [10.36040/jati.v9i6.16452](https://doi.org/10.36040/jati.v9i6.16452).