



Analisis Algoritma Fuzzy C-Means Untuk Pengelompokan Data Keluarga

Wahyu Cahyadi, Elin Haerani*, Alwis Nazir, Iwan Iskandar

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11950113467@students.uin-suska.ac.id, ^{2,*}elin.haerani@uin-suska.ac.id, ³alwis.nazir@uin-suska.ac.id, ⁴iwan.iskandar@uin-suska.ac.id,

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Abstrak—Pemetaan kondisi sosial ekonomi masyarakat berperan penting dalam mendukung perencanaan pembangunan yang tepat sasaran di tingkat desa. Penelitian ini bertujuan menerapkan algoritma Fuzzy C-Means (FCM) untuk mengelompokkan keluarga di Desa Bina Baru berdasarkan indikator sosial, ekonomi, dan kondisi lingkungan rumah tangga. Variabel yang digunakan meliputi jumlah anggota keluarga, sumber penghasilan, kondisi fisik rumah, fasilitas dasar, serta tingkat pengeluaran dan penghasilan bulanan. Penelitian ini menggunakan data kependudukan Desa Bina Baru yang terdiri atas 1.000 entri dengan 16 variabel. Algoritma FCM dipilih karena mampu mengakomodasi derajat keanggotaan ganda (fuzzy membership), sehingga lebih adaptif dalam menangkap keragaman dan ketidaktegasan karakteristik sosial ekonomi masyarakat. Hasil penelitian menunjukkan bahwa FCM menghasilkan dua cluster utama, yaitu Cluster 0 yang berjumlah 440 anggota mencerminkan keluarga dengan kondisi ekonomi menengah ke bawah, hunian permanen, dan fasilitas dasar yang memadai; serta Cluster 1 yang berjumlah 560 anggota menggambarkan keluarga dengan kondisi ekonomi lebih rendah, hunian semi permanen, dan jumlah anggota keluarga yang relatif lebih sedikit. Evaluasi menggunakan indeks Xie–Beni (35.4976), Fuzzy Partition Entropy (0.6843), dan Fuzzy Cluster Index (0.4468) menunjukkan bahwa model dua cluster memiliki kualitas pengelompokan terbaik dibandingkan jumlah cluster lainnya. Secara keseluruhan, algoritma Fuzzy C-Means efektif dalam memetakan variasi kesejahteraan keluarga dan dapat dimanfaatkan sebagai dasar dalam penyusunan kebijakan pembangunan serta program pemberdayaan masyarakat berbasis data di Desa Bina Baru.

Kata Kunci: Fuzzy C-Means; Xie-Beni; Fuzzy Partition Entropy; Fuzzy Cluster Index; Kesejahteraan Keluarga

Abstract—Mapping the socio-economic conditions of the community plays a crucial role in supporting targeted development planning at the village level. This study aims to apply the Fuzzy C-Means (FCM) algorithm to cluster families in Bina Baru Village based on social, economic, and household environmental indicators. The variables used include family size, income sources, physical condition of the house, basic facilities, as well as monthly expenditure and income levels. This study uses population data from Bina Baru Village, consisting of 1,000 entries with 16 variables. The FCM algorithm was chosen for its ability to accommodate multiple degrees of membership (fuzzy membership), making it more adaptable in capturing the diversity and ambiguity of socio-economic characteristics. The results show that FCM produces two main clusters: Cluster 0, with 440 members, reflects families with middle to lower economic conditions, permanent housing, and adequate basic facilities; and Cluster 1, with 560 members, represents families with lower economic conditions, semi-permanent housing, and relatively smaller family sizes. Evaluation using the Xie–Beni index (35.4976), Fuzzy Partition Entropy (0.6843), and Fuzzy Cluster Index (0.4468) indicates that the two-cluster model has the best clustering quality compared to other numbers of clusters. Overall, the Fuzzy C-Means algorithm is effective in mapping variations in family welfare and can be used as a basis for formulating development policies and data-driven community empowerment programs in Bina Baru Village.

Keywords: Fuzzy C-Means; Xie-Beni; Fuzzy Partition Entropy; Fuzzy Cluster Index; Family Welfare

1. PENDAHULUAN

Peningkatan kualitas hidup masyarakat menjadi salah satu fokus utama dalam pembangunan berkelanjutan. Salah satu langkah penting dalam mencapai tujuan tersebut adalah dengan melakukan pemetaan kondisi sosial ekonomi keluarga secara menyeluruh. Melalui pemetaan ini, pemerintah desa dan pihak terkait dapat memperoleh pemahaman yang lebih mendalam mengenai tingkat kesejahteraan masyarakat sehingga kebijakan yang dirancang dapat lebih tepat sasaran (Amelia et al., 2023). Desa Bina Baru merupakan salah satu wilayah pedesaan yang memiliki keragaman dalam kondisi sosial, ekonomi, dan hunian masyarakatnya. Kondisi ini menjadikan analisis terhadap faktor-faktor sosial ekonomi masyarakat desa tersebut sebagai hal yang penting untuk dilakukan, terutama untuk mendukung kebijakan pembangunan berbasis data (data-driven development) (Huddin et al., 2023; Kurniawati, 2022).

Namun, hingga saat ini proses pengelompokan keluarga berdasarkan kondisi sosial ekonomi di tingkat desa umumnya masih dilakukan secara manual dan subjektif. Pendekatan tersebut sering kali menimbulkan ketidaktepatan dalam menentukan kategori kesejahteraan keluarga, sehingga tidak jarang kebijakan bantuan sosial kurang tepat sasaran (Hidayat et al., 2023; Sanusi et al., 2020). Beberapa keluarga yang seharusnya termasuk dalam kelompok rentan justru tidak teridentifikasi secara akurat, sementara sebagian lainnya menerima bantuan meskipun tidak termasuk prioritas utama. Permasalahan ini menunjukkan perlunya sistem yang mampu mengelompokkan data masyarakat secara objektif dan berbasis pada parameter sosial ekonomi yang terukur. Oleh karena itu, penelitian ini mengajukan penerapan algoritma Fuzzy C-Means (FCM) sebagai metode yang dapat membantu proses klasifikasi keluarga berdasarkan tingkat kesejahteraan secara lebih terstruktur dan efisien (Faried, 2022; Muhammad & Budianita, 2022).

Penelitian juga berfokus pada bagaimana hasil pengelompokan tersebut dapat dimanfaatkan oleh pemerintah desa sebagai dasar pengambilan keputusan dalam penyusunan program bantuan dan kebijakan pemberdayaan masyarakat (Mustakim et al., 2024). Dengan demikian, penelitian ini diharapkan mampu memberikan solusi berbasis

data terhadap permasalahan dalam pengklasifikasian tingkat kesejahteraan masyarakat serta mendorong efisiensi dalam distribusi bantuan sosial (An-naziz Safaat et al., 2024; Surya Maulana et al., 2023).

Dalam konteks analisis sosial ekonomi, metode pengelompokan (clustering) memiliki peran penting dalam mengidentifikasi kelompok masyarakat berdasarkan kesamaan karakteristik. Salah satu metode clustering yang banyak digunakan karena fleksibilitasnya adalah Fuzzy C-Means (FCM). Metode ini memungkinkan setiap data untuk memiliki derajat keanggotaan pada lebih dari satu kelompok, sehingga lebih mampu merepresentasikan kondisi data yang bersifat tidak pasti atau kontinu (Novianti et al., 2022; Zahrani Putri et al., 2023). FCM telah banyak diterapkan dalam berbagai bidang, seperti segmentasi pelanggan (Zahrani Putri et al., 2023), pengelompokan provinsi berdasarkan penyakit menular (Novianti et al., 2022), hingga analisis data kesehatan (Faried, 2022), karena kemampuannya menghasilkan pengelompokan yang lebih halus dibandingkan algoritma clustering konvensional seperti K-Means.

Berdasarkan penelitian terdahulu, metode Fuzzy C-Means (FCM) telah banyak digunakan dalam pengelompokan data kemiskinan, salah satunya oleh Nurrahmah et al. (2020) yang menunjukkan efektivitas FCM dalam mengelompokkan data berdasarkan indikator sosial ekonomi (Kahar et al., 2019). Namun, penelitian tersebut masih bersifat umum dan belum mempertimbangkan karakteristik lokal secara spesifik pada tingkat desa. Penelitian ini berbeda karena menerapkan FCM pada data kemiskinan Desa Bina Baru dengan pemilihan variabel yang disesuaikan dengan kondisi lapangan, tanpa melakukan modifikasi algoritma. Dengan demikian, celah penelitian yang belum diselesaikan oleh penelitian sebelumnya adalah kurangnya kajian pengelompokan kemiskinan berbasis desa yang mampu merepresentasikan kondisi lokal secara lebih akurat, dan celah inilah yang berupaya diisi dalam penelitian ini.

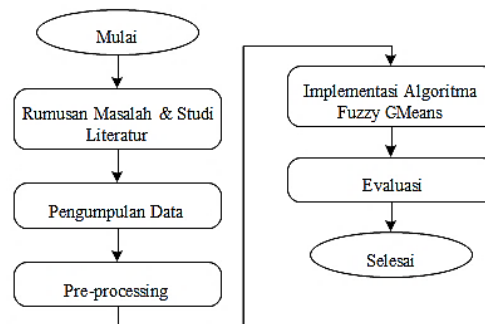
Data yang dikumpulkan dari Desa Bina Baru mencakup berbagai indikator seperti tingkat penghasilan, kondisi rumah, sumber air, jenis lantai, dan status kepemilikan rumah. Keberagaman variabel ini memungkinkan penerapan algoritma FCM untuk mengelompokkan keluarga berdasarkan kesamaan karakteristik sosial ekonomi. Melalui hasil pengelompokan tersebut, dapat diidentifikasi kelompok keluarga yang tergolong sejahtera maupun rentan, sehingga dapat menjadi dasar bagi pemerintah desa dalam merancang program bantuan atau kebijakan pemberdayaan masyarakat yang lebih efektif (Amelia et al., 2023; Kurniawati, 2022).

Penerapan algoritma Fuzzy C-Means dalam pengelompokan data sosial ekonomi memberikan nilai tambah dalam konteks pembangunan berbasis data (data-driven development). Dengan hasil pengelompokan yang akurat, proses pengambilan keputusan dapat dilakukan secara lebih objektif, efisien, dan berorientasi pada kebutuhan masyarakat (Amelia et al., 2023; Kahar et al., 2019; Mustakim et al., 2024). Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan algoritma Fuzzy C-Means dalam proses pengelompokan keluarga di Desa Bina Baru berdasarkan faktor sosial dan ekonomi yang ada. Hasil penelitian ini diharapkan dapat memberikan kontribusi nyata dalam penyusunan kebijakan pembangunan desa serta membantu pemerintah dalam menentukan penerima bantuan yang tepat sasaran.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian merupakan rangkaian proses yang ditempuh peneliti untuk melaksanakan penelitian secara teratur dan terorganisir. Alur tahapan penelitian dalam studi ini dapat dilihat pada Gambar 1 berikut:



Gambar 1. Tahapan penelitian

Gambar 1 memperlihatkan tahapan penelitian yang dilakukan, mulai dari perumusan masalah dan studi literatur, pengumpulan data, pra-pemrosesan, implementasi algoritma Fuzzy C-Means, hingga tahap evaluasi.

2.2 Pengumpulan Data

Penelitian ini memanfaatkan dataset sekunder yang bersumber dari studi terdahulu yang dilakukan oleh Felina Amelia (2022) berjudul “Clustering Keluarga Miskin Desa Bina Baru dengan Metode K-Medoids”, yang menggunakan data kependudukan Desa Bina Baru sebanyak 1.000 entri dengan 16 variabel. Dataset ini masih relevan digunakan kembali karena mencerminkan kondisi sosial ekonomi masyarakat desa yang relatif stabil dalam kurun waktu penelitian serta belum mengalami perubahan signifikan pada struktur indikator yang digunakan. Selain itu, data tersebut dikumpulkan



menggunakan teknik sensus, yaitu mencakup seluruh data keluarga yang tercatat pada wilayah penelitian, sehingga memiliki tingkat representativitas yang tinggi dan layak digunakan sebagai dasar analisis dalam penelitian ini.

2.3 Pre-processing

Pada tahapan ini dilakukan serangkaian proses untuk mengubah data awal menjadi bentuk yang lebih sederhana dan siap dianalisis.

a. Data Selection

Data Selection atau seleksi data merupakan tahapan yang diperlukan untuk menyeleksi atribut yang akan digunakan dalam penelitian (Sholeh et al., 2024). Dalam penelitian ini, atribut yang akan digunakan adalah Penghasilan, Pengeluaran, Anggota Keluarga, Sumber Penghasilan, Dinding Rumah, Lantai Rumah, Status Rumah, Atap Rumah, Tempat Buang Air Besar, MCK, Sumber Air Minum, Penerangan Utama, Bahan Bakar Memasak, Pembuangan Sampah, Pembuangan Limbah Cair, dan Kondisi Rumah.

b. Data Cleaning

Tahapan ini merupakan tahapan pembersihan data dari data yang missing value (Sharifnia et al., 2026). Tahapan ini dilakukan untuk mendapatkan hasil yang valid dari hasil clustering penelitian.

c. Data Transformation

Tahapan ini merupakan merubah data kebentuk yang diperlukan sesuai dengan pembahasan penelitian. Data yang telah di dibersihkan dari data-data yang tidak diperlukan kemudian akan berlanjut ke proses perubahan. Proses perubahan data yaitu mengubah data ke dalam bentuk angka menggunakan Label Encoding (Kumar & Bhardwaj, 2025).

2.4 Data Mining

Data mining merupakan proses analisis yang digunakan untuk mengekstraksi pola, tren, dan informasi berharga dari kumpulan data dalam skala besar (Srirahayu & Pribadie, 2023). Melalui penerapan teknik statistik, pembelajaran mesin, dan metode analisis lainnya, data mining mampu mengungkap wawasan yang tidak mudah ditemukan dengan analisis konvensional. Proses ini menjadi bagian penting dari Knowledge Discovery in Databases (KDD), yang mencakup tahapan mulai dari pembersihan, integrasi, seleksi, hingga transformasi data sebelum diterapkan algoritma analisis. Dalam berbagai bidang, seperti keuangan, kesehatan, pendidikan, maupun pemasaran, data mining telah terbukti bermanfaat dalam mendeteksi anomali, memahami perilaku, serta mendukung pengambilan keputusan berbasis data (An-nazir Safaat et al., 2024; Kahar et al., 2019).

2.5 Clustering

Clustering merupakan teknik analisis data yang digunakan untuk mengelompokkan sekumpulan objek berdasarkan kemiripan karakteristik tertentu. Tujuan utama dari metode ini adalah membagi data ke dalam kelompok (cluster) sehingga objek dalam satu cluster memiliki tingkat kesamaan yang tinggi, sementara objek antar cluster berbeda secara signifikan (Alasali & Ortakci, 2024). Berbeda dengan klasifikasi, clustering tidak membutuhkan label atau kategori awal sehingga bermanfaat untuk menemukan pola maupun struktur tersembunyi dalam data. Pendekatan ini banyak diterapkan di berbagai bidang, seperti pengenalan pola, pemasaran, analisis citra, hingga keamanan siber. Secara umum, metode clustering terbagi menjadi dua kategori, yaitu hierarkis, yang membentuk struktur pohon, dan non-hierarkis, yang mengelompokkan data tanpa struktur pohon tertentu (Mustakim et al., 2024). Beberapa metode Clustering yang populer antara lain centroid-based clustering, density-based clustering, dan distribution-based clustering. Metode centroid-based, seperti Fuzzy C-Means, menggunakan pusat cluster (centroid) yang diperbarui secara iteratif berdasarkan jarak data (Huddin et al., 2023). Sementara itu, density-based clustering, seperti DBSCAN, membentuk cluster berdasarkan kepadatan titik data di sekitarnya dan efektif dalam menangani outlier (Novianti et al., 2022). Adapun distribution-based clustering bekerja dengan asumsi bahwa data mengikuti distribusi probabilitas tertentu, misalnya menggunakan algoritma Expectation-Maximization (EM) untuk memperkirakan parameter distribusi dan mengelompokkan data sesuai probabilitasnya (Novianti et al., 2022). Dengan ragam pendekatan ini, clustering dapat memberikan hasil yang sesuai dengan kebutuhan analisis dan karakteristik data yang diteliti..

2.6 Algoritma Fuzzy C-Means

Algoritma Fuzzy C-Means (FCM) merupakan metode clustering yang memungkinkan suatu data memiliki keanggotaan pada lebih dari satu cluster dengan derajat keanggotaan tertentu yang dinyatakan dalam nilai antara 0 dan 1. Berbeda dengan hard clustering seperti K-Means, yang mengharuskan setiap data hanya masuk ke satu cluster secara eksklusif, FCM memberikan fleksibilitas dalam pengelompokan data yang bersifat tumpang tindih (overlapping). Dengan demikian, metode ini efektif digunakan untuk menganalisis data kompleks yang tidak memiliki batasan cluster yang

(Faried, 2022; Mustakim et al., 2024).
Proses algoritma FCM diawali dengan menentukan jumlah cluster (c) dan parameter fuzziness (m), yang mengontrol tingkat kekaburan antar cluster. Selanjutnya, setiap data diberikan nilai keanggotaan awal secara acak. Pusat cluster (centroid) kemudian dihitung menggunakan rata-rata tertimbang dari data dengan bobot berdasarkan derajat keanggotaan. Rumus centroid ditunjukkan sebagai berikut:



$$v_j = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m} \quad (1)$$

Di mana v_j adalah centroid dari cluster ke- j . u_{ij} adalah derajat keanggotaan data ke- i terhadap cluster ke- j . x_i adalah data ke- i dan m adalah parameter fuzziness. Setelah itu, derajat keanggotaan diperbarui berdasarkan jarak antara data dan centroid cluster. Rumus derajat keanggotaan diberikan sebagai berikut:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d(x_i, v_j)}{d(x_i, v_k)} \right)^{\frac{2}{m-1}}} \quad (2)$$

Di mana u_{ij} adalah derajat keanggotaan data ke- i terhadap cluster ke- j dan $d(x_i, v_j)$ adalah jarak antara data ke- i dan pusat cluster ke- j . Kemudian $d(x_i, v_k)$ adalah jarak antara data ke- i dan pusat cluster ke- k . m adalah parameter fuzziness ($m > 1$) yang mengontrol tingkat kekaburan batas antar cluster dan c adalah jumlah cluster. Proses ini dilakukan secara iteratif, yaitu menghitung centroid baru dan memperbarui nilai keanggotaan hingga perubahan nilai keanggotaan atau centroid berada di bawah ambang batas konvergensi. Hasil akhir FCM berupa matriks keanggotaan yang menunjukkan derajat keterikatan setiap data terhadap masing-masing cluster. Data dapat dipetakan ke cluster dengan nilai keanggotaan tertinggi, atau tetap dianalisis sebagai anggota parsial dari beberapa cluster (Surya Maulana et al., 2023).

2.7 Xie-Beni

Indeks Xie-Beni (XB) merupakan salah satu indeks validitas internal yang paling banyak digunakan dalam proses evaluasi hasil clustering fuzzy. Indeks ini diperkenalkan oleh Xie dan Beni (1991) sebagai ukuran yang mempertimbangkan dua aspek penting, yaitu kompaksi (compactness) dan pemisahan cluster (separation) dalam partisi fuzzy. Kompaksi menunjukkan seberapa rapat objek-objek data berkumpul di sekitar pusat cluster, sedangkan separation mengukur seberapa jauh antar pusat cluster satu dengan lainnya. Secara umum, nilai indeks Xie-Beni dirumuskan sebagai berikut:

$$XB = \frac{\sum_{i=1}^n \sum_{k=1}^c \mu_{ik}^m \|x_i - v_k\|^2}{n \cdot \min_{j \neq k} \|v_j - v_k\|^2} \quad (3)$$

Indeks XB memiliki sifat bahwa semakin kecil nilai XB, semakin baik kualitas pengelompokan yang dihasilkan, karena hal tersebut menunjukkan bahwa cluster yang terbentuk memiliki penyebaran internal yang kecil dan jarak antar pusat cluster yang besar (Fazel Zarandi et al., 2021).

2.8 Fuzzy Partition Entropy (FPE)

Fuzzy Partition Entropy (FPE) diperkenalkan sebagai indeks validitas yang mengukur tingkat ketidakpastian (uncertainty) dalam matriks partisi fuzzy. Suatu partisi dikatakan baik apabila sistem memiliki tingkat entropi yang rendah, artinya setiap data memiliki kecenderungan kuat untuk berada pada satu cluster tertentu. Sebaliknya, jika entropi tinggi, maka derajat keanggotaan data terhadap beberapa cluster cenderung merata sehingga cluster menjadi kurang jelas (Fazel Zarandi et al., 2021). Rumus umum FPE adalah:

$$FPE = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c \mu_{ik} \log(\mu_{ik}) \quad (4)$$

Nilai FPE berada pada rentang 0 hingga $\log(c)$ dan Nilai FPE kecil \rightarrow clustering baik, karena data memiliki keanggotaan dominan pada cluster tertentu. Nilai FPE besar \rightarrow clustering buruk, karena data tidak memiliki kecenderungan yang jelas pada cluster tertentu. FPE banyak digunakan pada metode FCM maupun metode fuzzy lainnya. Karena indeks ini mempertimbangkan derajat keanggotaan secara langsung, ia sangat sensitif terhadap penyebaran fuzzy membership. Hal ini membuat FPE cocok digunakan ketika partisi cluster bersifat ambigu atau ketika data mengandung noise.

2.9 Fuzzy Cluster Index (FCI)

Fuzzy Partition Cluster Index (FPCI), atau sering juga disebut Fuzzy Cluster Index (FCI), merupakan indeks validitas fuzzy clustering yang memadukan komponen kompaksi dan pemisahan cluster, mirip dengan Xie-Beni namun dengan formulasi berbeda. FPCI dikembangkan untuk meningkatkan kemampuan indeks dalam mendeteksi kualitas cluster pada kondisi fuzzy di mana derajat keanggotaan tidak bersifat mutlak. FPCI didefinisikan sebagai:

$$FCI = \frac{\sum_{i=1}^n \sum_{k=1}^c \mu_{ik}^m \|x_i - v_k\|^2}{\sum_{i=1}^n \sum_{k=1, j \neq k}^c \|v_j - v_k\|^2} \quad (5)$$

Nilai FCI, semakin baik kualitas cluster. Nilai kecil menunjukkan penyebaran intra-cluster yang kompak serta jarak antar pusat cluster yang besar. FPCI termasuk fungsi indeks berbasis jarak dan sangat cocok digunakan untuk algoritma fuzzy seperti FCM, PCM, FPCM, hingga PFCM. Penggunaan indeks ini terbukti mampu memvalidasi cluster pada data yang mengandung noise, karena ia mempertimbangkan faktor fuzzy membership secara eksplisit. Dalam

praktik penelitian modern, FPCI sering digunakan bersama dengan indeks lain seperti Xie–Beni dan Partition Entropy untuk mendapatkan validitas menyeluruh (Rathee et al., 2022).

3. HASIL DAN PEMBAHASAN

3.1 Penentuan Cluster Terbaik

Melalui serangkaian pengujian, diperoleh nilai yang menunjukkan bahwa konfigurasi dua cluster memberikan pemisahan yang relatif lebih baik dibandingkan opsi lainnya. Dengan demikian, jumlah cluster terbaik yang digunakan dalam penelitian ini ditetapkan sebanyak dua, karena mampu menghasilkan struktur pengelompokan yang paling mendekati pola alami dalam data.

```
=== Evaluasi Clustering FCM (XB, FPE, FCI) ===
Cluster=2, XB=35.4976, FPE=0.6843, FCI=0.4468
Cluster=3, XB=380.9669, FPE=1.0813, FCI=0.6064
Cluster=4, XB=31766.1591, FPE=1.3521, FCI=0.6822
Cluster=5, XB=266759.9539, FPE=1.5559, FCI=0.7246
Cluster=6, XB=8339686.6094, FPE=1.7233, FCI=0.7563
Cluster=7, XB=9157892047.6111, FPE=1.8665, FCI=0.7804
Cluster=8, XB=2820175.7608, FPE=1.9918, FCI=0.8005
Cluster=9, XB=108659837416191.8594, FPE=2.0915, FCI=0.7935
Cluster=10, XB=78877098.4266, FPE=2.2033, FCI=0.8283
```

Gambar 2. Hasil Evaluasi Xie-Beni, FPE dan FCI untuk Penentuan Jumlah Cluster Optimum

Evaluasi kualitas cluster pada penelitian ini dilakukan menggunakan tiga indeks validitas fuzzy, yaitu Xie–Beni (XB), Fuzzy Partition Entropy (FPE), dan Fuzzy Cluster Index (FCI). Ketiga indeks ini berfungsi mengevaluasi apakah jumlah cluster yang dibentuk Fuzzy C-Means (FCM) sudah menghasilkan pengelompokan yang kompak, terpisah dengan baik, dan memiliki kepastian fuzzy yang memadai. Secara umum, nilai terbaik ditandai dengan:

XB paling kecil → cluster semakin kompak dan terpisah.

FPE paling kecil → ketidakpastian partisi rendah.

FCI paling kecil → keseimbangan compactness–separation optimal.

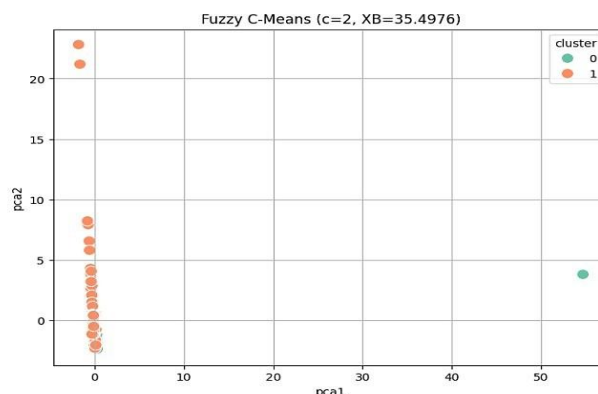
3.2 Penerapan Algoritma Fuzzy C-Means Menggunakan Python

Pada tahap ini, algoritma Fuzzy C-Means diimplementasikan menggunakan Python untuk melakukan pengelompokan berdasarkan karakteristik data yang tersedia. Proses penentuan jumlah cluster tidak ditentukan secara objektif, melainkan melalui evaluasi menggunakan Xie-Beni, FPE dan FCI untuk menilai seberapa baik setiap data berada dalam cluster yang sesuai dibandingkan dengan cluster lainnya. Setelah jumlah cluster optimal diperoleh, tahap selanjutnya adalah melakukan normalisasi data menggunakan MinMaxScaler agar setiap atribut berada pada skala yang sama dan tidak menimbulkan bias dalam proses perhitungan. Pada algoritma Fuzzy C-Means, proses awal dilakukan dengan memberikan derajat keanggotaan awal secara acak untuk setiap data terhadap seluruh cluster. Centroid kemudian dihitung sebagai rata-rata tertimbang berdasarkan nilai keanggotaan tersebut. Proses pembaruan nilai keanggotaan dan perhitungan centroid dilakukan secara iteratif hingga mencapai kondisi konvergen, sehingga data dapat dikelompokkan secara lebih fleksibel sesuai tingkat kedekatannya dengan masing-masing cluster.

Tabel 1. Jumlah Anggota Tiap Cluster

Cluster	Jumlah
Cluster 0	440
Cluster 1	560

Pada Tabel 1, jumlah anggota tiap cluster di dapat nilai cluster 0 sebanyak 440 dan cluster 1 sebanyak 560.



Gambar 3. Persebaran Data Hasil Iterasi FCM



Gambar 3 menunjukkan hasil pengelompokan data menggunakan algoritma Fuzzy C-Means dengan jumlah cluster optimal sebanyak dua. Visualisasi dilakukan menggunakan reduksi dimensi Principal Component Analysis (PCA) agar struktur data yang sebenarnya berdimensi tinggi dapat ditampilkan secara lebih sederhana dalam bentuk dua dimensi. Pendekatan ini memudahkan peneliti untuk melihat pola penyebaran data secara keseluruhan, termasuk bagaimana data saling berdekatan atau saling menjauh. Pada grafik tersebut, setiap titik mewakili satu entitas data, dan perbedaan warna digunakan untuk menunjukkan keanggotaan cluster yang terbentuk. Dengan demikian, pembaca dapat secara langsung mengamati hasil pemetaan dan mendapatkan gambaran intuitif mengenai struktur kelompok dalam dataset.

Dari visualisasi tersebut terlihat bahwa sebagian besar data terkonsentrasi pada Cluster 1 (warna oranye), yang menunjukkan adanya kelompok dengan karakteristik dominan dan lebih homogen. Sebaliknya, hanya sebagian kecil data yang berada pada Cluster 0 (warna hijau), sehingga cluster ini tampak berisi data yang memiliki karakteristik berbeda atau menyimpang dari pola umum kelompok utama. Penyebaran ini mengindikasikan bahwa mayoritas rumah tangga dalam dataset memiliki kesamaan kondisi yang cukup kuat, sementara sebagian lainnya membentuk kelompok kecil dengan ciri khas tertentu. Kehadiran cluster minor ini memberi gambaran awal bahwa terdapat variasi yang signifikan namun tidak terlalu besar di antara responden, yang penting untuk dianalisis lebih lanjut pada tahap interpretasi karakteristik cluster.

3.3 Centroid Hasil Pengujian

Pada tahapan ini, hasil temuan atau pengetahuan yang ditemukan pada dataset disajikan menggunakan metode algoritma Fuzzy C-Means. Dijelaskan perumusan keputusan atau tindakan berdasarkan hasil yang diperoleh.

Tabel 2. Nilai Centroid

Variabel	Cluster 0	Cluster 1
Anggota keluarga	3,527273	2,883929
Sumber penghasilan	1,756818	1,419643
Dinding rumah	1,054545	1,2
Lantai rumah	1,711364	1,351786
Status rumah	1,115909	1,203571
Atap rumah	1,072727	2
Tempat buang air besar	1	1,001786
Mck	1	1,001786
Sumber air minum	1,052273	1,023214
Penerangan utama	1	1,001786
Bahan bakar memasak	1,038636	1,008929
Pembuangan sampah	1,077273	1,069643
Pembuangan limbah cair	1,040909	1,030357
Kondisi rumah	1,993182	2
Penghasilan	4673523	2685196
Pengeluaran	3529091	1852321

Untuk memberikan gambaran yang lebih nyata mengenai hasil klasterisasi, Tabel 2 menampilkan titik centroid dari setiap cluster diubah kembali ke dalam bentuk data asli yang mudah dikenali dalam konteks kehidupan masyarakat. Langkah ini dilakukan karena angka-angka pada hasil perhitungan sering kali sulit menggambarkan kondisi sebenarnya di lapangan. Dengan menerjemahkan nilai centroid ke bentuk data yang lebih konkret, hasil analisis menjadi lebih dekat dengan situasi yang dialami rumah tangga sehari-hari. Proses ini juga membantu memperlihatkan pola-pola penting yang mungkin tidak terlihat jika hanya disajikan dalam bentuk statistik semata.

Melalui pendekatan ini, karakteristik tiap kelompok dapat dipahami tidak hanya sebagai rangkaian nilai, tetapi sebagai cerminan nyata dari kehidupan keluarga di masing-masing cluster. Misalnya, perbedaan pada jumlah anggota keluarga, tingkat pengeluaran, atau kondisi hunian dapat terlihat lebih jelas ketika disajikan kembali dalam bentuk data asli. Dengan demikian, hasil klasterisasi menjadi lebih mudah diinterpretasikan, lebih relevan bagi pengambil keputusan, dan lebih bermanfaat dalam menggambarkan variasi kebutuhan serta tantangan yang dihadapi oleh masyarakat di berbagai kelompok.

Tabel 3. Karakteristik Tiap Cluster

Variabel	Cluster 0 (Ekonomi Sejahtera)	Cluster 1 (Ekonomi Pra-Sejahtera)
Anggota keluarga	3-4 orang	2-3 orang
Sumber penghasilan	Buruh tani – Pemilik Kebun Sawit	Pemilik Kebun Sawit – Pedagang
Dinding rumah	Semen	Semen – Kayu
Lantai rumah	Keramik – Kayu	Semen-Keramik
Status rumah	Sendiri	Sendiri-Kontrak
Atap rumah	Seng	Genteng - Seng
Tempat buang air besar	Jamban sendiri	Jamban sendiri



Variabel	Cluster 0 (Ekonomi Sejahtera)	Cluster 1 (Ekonomi Pra-Sejahtera)
MCK	Sendiri	Sendiri
Sumber air minum	Air isi ulang	Air isi ulang – Mata air/sumur
Penerangan utama	Listrik PLN	Listrik PLN
Bahan bakar memasak	Gas kota/LPG	Gas kota/LPG – Minyak Tanah
Pembuangan sampah	Dibakar – Tempat sampah	Dibakar – Tempat sampah
Pembuangan limbah cair	Lubang di tanah – Tangki	Lubang di tanah – Tangki
Kondisi rumah	Tidak kumuh	Tidak kumuh
Penghasilan	± Rp 4.673.523 / bulan	± Rp 2.685.196 / bulan
Pengeluaran	± Rp 3.529.091 / bulan	± Rp 1.852.321 / bulan

Tabel 3 menyajikan hasil pengelompokan data keluarga Desa Bina Baru menggunakan algoritma Fuzzy C-Means (FCM). Pengelompokan dilakukan berdasarkan variabel kondisi sosial, ekonomi, dan lingkungan rumah tangga, yang mencakup jumlah anggota keluarga, sumber penghasilan, kondisi rumah, akses terhadap fasilitas dasar, hingga tingkat penghasilan dan pengeluaran bulanan. Dari hasil clustering, diperoleh dua kelompok utama keluarga yang memiliki karakteristik berbeda. Berikut adalah analisis karakteristik dari setiap cluster:

- Cluster 0 merepresentasikan kelompok keluarga dengan kondisi ekonomi menengah bawah yang relatif stabil. Keluarga dalam cluster ini umumnya memiliki 3 hingga 4 anggota keluarga, menunjukkan struktur keluarga inti yang lengkap. Sumber penghasilan mereka berasal dari sektor pertanian, terutama sebagai buruh tani dan pemilik kebun sawit, yang mencerminkan adanya kombinasi antara pekerjaan harian dan kepemilikan aset produktif. Secara fisik, rumah mereka cenderung permanen dengan dinding semen dan lantai keramik atau kayu, serta berstatus milik sendiri, menandakan kestabilan tempat tinggal. Fasilitas dasar seperti MCK dan jamban dimiliki sendiri, sumber air minum berasal dari air isi ulang, dan penerangan utama menggunakan listrik PLN, menunjukkan bahwa kebutuhan dasar telah terpenuhi dengan baik. Bahan bakar memasak yang digunakan adalah gas kota atau LPG, yang umum digunakan oleh masyarakat dengan tingkat ekonomi menengah. Sistem pembuangan sampah dilakukan dengan dibakar atau menggunakan tempat sampah, sementara limbah cair dibuang ke lubang tanah atau tangki. Rumah mereka tergolong tidak kumuh, menandakan lingkungan yang cukup tertata. Dengan penghasilan rata-rata Rp 4.673.523 per bulan dan pengeluaran sekitar Rp 3.529.091 per bulan, kelompok ini masih memiliki sisa pendapatan, sehingga dapat dikatakan bahwa mereka memiliki tingkat kesejahteraan yang cukup baik dibandingkan cluster lainnya..
- Cluster 1 menggambarkan kelompok keluarga dengan kondisi ekonomi lebih rendah dibandingkan Cluster 0. Jumlah anggota keluarga cenderung lebih sedikit, yakni 2 hingga 3 orang, yang menunjukkan kemungkinan keluarga muda atau keluarga kecil dengan tanggungan terbatas. Sumber penghasilan utama berasal dari pemilik kebun sawit berskala kecil dan pedagang, yang menunjukkan orientasi ekonomi pada sektor informal dan usaha mikro. Kondisi fisik rumah dalam cluster ini umumnya semi permanen, dengan dinding semen dan kayu, serta lantai yang bervariasi antara semen dan keramik. Status kepemilikan rumah bervariasi antara milik sendiri dan kontrak, menandakan sebagian anggota belum memiliki hunian tetap. Material atap yang digunakan adalah genteng atau seng, menunjukkan kualitas bangunan yang beragam. Meskipun demikian, fasilitas sanitasi cukup baik karena mereka memiliki jamban dan MCK sendiri, serta menggunakan air isi ulang atau air sumur/mata air sebagai sumber air minum. Penerangan utama tetap menggunakan listrik PLN, dan bahan bakar memasak yang digunakan adalah kombinasi antara gas LPG dan minyak tanah, menunjukkan adanya variasi sesuai kemampuan ekonomi. Sistem pembuangan sampah dan limbah cair serupa dengan Cluster 0, yaitu dengan pembakaran dan saluran sederhana. Rumah-rumah dalam cluster ini tetap tergolong tidak kumuh, meskipun sederhana. Dengan penghasilan rata-rata Rp 2.685.196 per bulan dan pengeluaran sekitar Rp 1.852.321 per bulan, kelompok ini menunjukkan daya beli yang lebih rendah serta ruang keuangan yang lebih sempit, sehingga dapat dikategorikan sebagai keluarga dengan taraf hidup sederhana dan ekonomi rentan.

Berdasarkan hasil analisis kedua cluster, dapat disimpulkan bahwa terdapat perbedaan yang cukup jelas antara Cluster 0 dan Cluster 1 dalam hal kondisi sosial ekonomi dan karakteristik tempat tinggal. Cluster 0 yang berjumlah 440 anggota menunjukkan kelompok keluarga dengan kondisi ekonomi lebih stabil dan sejahtera, ditandai oleh pendapatan yang lebih tinggi, kepemilikan rumah sendiri, serta penggunaan fasilitas dasar yang lebih baik seperti lantai keramik, air isi ulang, dan bahan bakar LPG. Rumah mereka umumnya permanen dengan sanitasi yang memadai, sehingga dapat dikategorikan sebagai keluarga menengah bawah yang mapan. Sebaliknya, Cluster 1 yang berjumlah 560 anggota menggambarkan kelompok keluarga dengan taraf ekonomi yang lebih rendah, ditandai oleh pendapatan yang lebih kecil, beberapa di antaranya masih menyewa rumah, serta kondisi hunian semi permanen dengan fasilitas dasar yang sederhana. Meskipun demikian, kelompok ini tetap memiliki akses terhadap listrik, air bersih, dan fasilitas sanitasi, menunjukkan bahwa taraf hidup mereka masih tergolong layak namun rentan secara ekonomi. Secara keseluruhan, hasil klusterisasi ini menunjukkan adanya tingkatan kesejahteraan yang berbeda di antara dua kelompok masyarakat, di mana perbedaan utama terletak pada tingkat pendapatan, kepemilikan aset, dan kualitas tempat tinggal.

3.4 Evaluasi

Berdasarkan Gambar 3 diperoleh analisa hasil evaluasi sebagai berikut.



1. Analisis Indeks Xie–Beni (XB)

Hasil perhitungan menunjukkan pola yang sangat jelas bahwa nilai XB terkecil berada pada jumlah cluster = 2, yaitu sebesar 35.4976. Sementara itu, pada jumlah cluster yang lebih besar (3 hingga 10 cluster), nilai XB meningkat secara drastis hingga mencapai skala jutaan dan bahkan triliunan. Peningkatan ekstrem ini menandakan bahwa semakin banyak cluster yang dibentuk, kualitas pemisahan antar cluster justru memburuk, ditandai dengan jarak antar pusat cluster (separation) yang mengecil atau penyebaran antar data dalam cluster meningkat. Dengan demikian, dari sudut pandang XB, cluster terbaik adalah 2 cluster.

2. Analisis Indeks Fuzzy Partition Entropy (FPE)

Indeks FPE juga memperlihatkan kecenderungan yang konsisten. Nilai FPE meningkat seiring bertambahnya jumlah cluster, mulai dari 0.6843 pada 2 cluster hingga 2.2033 pada 10 cluster. FPE yang rendah menandakan bahwa derajat keanggotaan setiap data terhadap cluster dominan lebih tegas dan tidak ambigu. Oleh karena itu, 2 cluster kembali memberikan hasil partisi paling jelas dan paling sedikit ketidakpastian, dibandingkan jumlah cluster lainnya yang menunjukkan entropi lebih tinggi.

3. Analisis Indeks Fuzzy Cluster Index (FCI)

FCI juga menunjukkan kecenderungan peningkatan nilai ketika jumlah cluster bertambah. Nilai FCI terkecil terdapat pada cluster = 2, yaitu 0.4468, kemudian meningkat bertahap hingga mencapai 0.8283 pada 10 cluster. Karena nilai FCI yang lebih kecil menunjukkan kualitas cluster yang lebih baik (kompak dan saling terpisah), maka kembali dapat disimpulkan bahwa jumlah cluster yang optimal adalah 2.

3.5 Perbandingan dengan Penelitian Sebelumnya

Penelitian ini menggunakan algoritma Fuzzy C-Means (FCM) untuk mengelompokkan data keluarga di Desa Bina Baru dan menghasilkan dua cluster utama yang mencerminkan perbedaan kondisi sosial ekonomi keluarga. Hasil ini berbeda dengan penelitian yang dilakukan oleh Amelia et al. yang menggunakan algoritma K-Medoids terhadap data Desa Bina Baru tahun 2020 dan menghasilkan tiga cluster, yaitu kelompok kaya, menengah, dan miskin (Amelia et al., 2023). Perbedaan jumlah cluster tersebut menunjukkan bahwa K-Medoids menghasilkan pengelompokan yang lebih tegas dengan pembagian kategori yang jelas, sementara FCM cenderung menyederhanakan pembagian kluster dengan menekankan pemisahan kondisi ekonomi utama berdasarkan pola data. Secara metodologis, K-Medoids menerapkan pendekatan hard clustering sehingga setiap keluarga hanya dapat menjadi anggota satu cluster tertentu, sedangkan FCM menggunakan pendekatan soft clustering dengan derajat keanggotaan yang memungkinkan satu data memiliki kedekatan dengan lebih dari satu cluster. Pendekatan ini menjadikan FCM mampu memberikan hasil pengelompokan yang lebih halus dan fleksibel, terutama dalam merepresentasikan keluarga dengan kondisi sosial ekonomi yang berada pada batas antar cluster. Dengan demikian, meskipun K-Medoids lebih mudah diinterpretasikan karena pembagiannya bersifat tegas, FCM dinilai lebih adaptif dan informatif untuk analisis kemiskinan yang bersifat kompleks dan multidimensi.

4. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma Fuzzy C-Means (FCM) mampu mengelompokkan data keluarga secara efektif berdasarkan indikator sosial, ekonomi, dan kondisi lingkungan rumah tangga. Proses klusterisasi menghasilkan dua kelompok utama yang mencerminkan perbedaan kondisi sosial ekonomi keluarga di Desa Bina Baru, di mana Cluster 0 menggambarkan keluarga dengan kondisi ekonomi menengah ke bawah, rata-rata 3–4 anggota keluarga, sumber penghasilan utama dari sektor pertanian, serta hunian permanen dengan fasilitas dasar yang memadai, sedangkan Cluster 1 merepresentasikan keluarga dengan kondisi ekonomi lebih rendah, jumlah anggota keluarga lebih sedikit, hunian semi permanen dengan fasilitas terbatas, dan sumber penghasilan yang lebih beragam. Evaluasi kualitas klusterisasi menggunakan indeks Xie–Beni (XB), Fuzzy Partition Entropy (FPE), dan Fuzzy Cluster Index (FCI) menunjukkan bahwa model dengan 2 cluster memberikan performa terbaik dengan nilai $XB = 35.4976$, $FPE = 0.6843$, dan $FCI = 0.4468$, yang menandakan tingkat kompaksi dan pemisahan cluster yang optimal serta ketidakpastian partisi yang rendah, sementara peningkatan jumlah cluster justru menurunkan kualitas pengelompokan. Secara umum, pendekatan FCM yang digunakan dalam penelitian ini tidak terbatas pada Desa Bina Baru dan berpotensi diterapkan pada desa lain di Indonesia atau pada dataset kemiskinan dengan karakteristik serupa untuk membantu pemetaan kondisi sosial ekonomi masyarakat. Namun demikian, penelitian ini masih memiliki keterbatasan, antara lain penggunaan dataset sekunder yang belum mencerminkan perubahan kondisi terbaru serta belum adanya perbandingan dengan metode klusterisasi lain, sehingga penelitian selanjutnya disarankan menggunakan data yang lebih mutakhir, membandingkan FCM dengan algoritma lain, serta mengintegrasikan hasil klusterisasi dengan indikator kebijakan untuk meningkatkan relevansi dan akurasi pengambilan keputusan.

REFERENCES

- Alasali, T., & Ortakci, Y. (2024). Clustering Techniques in Data Mining: A Survey of Methods, Challenges, and Applications. *Computer Science*. <https://doi.org/10.53070/bbd.1421527>
- Amelia, F., Iskandar, I., Kurnia Gusti, S., & Haerani, E. (2023). Clustering Keluarga Miskin Desa Bina Baru Dengan Metode K-Medoids. *Krea-Tif: Jurnal Teknik Informatika*, 11(1), 1–13. <https://doi.org/10.32832/krea-tif.v11i1.14104>



- An-naziz Safaat, W., Kurniawan, R., & Arie Wijaya, Y. (2024). Penerapan Data Mining Clustering Menggunakan Algoritma Fuzzy C-Means Pada Data Pencari Kerja Di Kabupaten Kuningan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1507–1511. <https://doi.org/10.36040/jati.v8i2.8411>
- Faried, M. A. F. (2022). *Pengelompokan data kelas BPJS Kesehatan menggunakan metode Fuzzy C-Means (FCM)*. Universitas Islam Negeri Sultan Syarif Kasim Riau .
- Fazel Zarandi, M. H., Sotodian, S., & Castillo, O. (2021). A New Validity Index for Fuzzy-Possibilistic C-Means Clustering. *Scientia Iranica*, 0(0), 0–0. <https://doi.org/10.24200/sci.2021.50287.1614>
- Hidayat, Y., Nazir, A., Candra, R. M., Sanjaya, S., & Syafria, F. (2023). Clustering Vaksinasi Penyakit Mulut dan Kuku Menggunakan Algoritma Fuzzy C-Means. *Journal of Computer System and Informatics (JoSYC)*, 4(3), 587–593. <https://doi.org/10.47065/josyc.v4i3.3416>
- Huddin, S., Haerani, E., & Oktavia, L. (2023). Penerapan Fuzzy C-Means Pada Klusterisasi Penerima Bantuan Pangan Non Tunai. *Media Online*, 4(1), 453–461. <https://doi.org/10.30865/klik.v4i1.988>
- Kahar, N. F., Hadjaratie, L., Suhada, S., & Padiku, I. R. (2019). Implementasi Data Mining Dalam Penentuan Tingkat Kemiskinan Menggunakan Fuzzy C-Means. *Jambura Journal of Informatics*, 1(1), 27–36. <https://doi.org/10.37905/jji.v1i1.2332>
- Kumar, M., & Bhardwaj, V. (2025). Evaluating Label Encoding and Preprocessing Techniques for Breast Cancer Prediction Using Machine Learning Algorithms. *International Journal of Computational Intelligence Systems*, 18(1), 218. <https://doi.org/10.1007/s44196-025-00957-7>
- Kurniawati, V. (2022). *Analisis clustering kemiskinan di Indonesia menggunakan algoritma Fuzzy C-Means dan Fuzzy Possibilistic C-Means*. UNIVERSITAS LAMPUNG.
- Muhammad, A., & Budianita, E. (2022). Pengelompokan Tingkat Kecanduan Game Online Menggunakan Algoritma Fuzzy C-Means. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 5(4), 601–610. <https://doi.org/10.32672/jnkti.v5i4.4511>
- Mustakim, M., Aini, D. N., Batubara, A. U., Erkamim, Moh., & Legito, L. (2024). Fuzzy Clustering-Based Grouping for Mapping the Distribution of Student Success Data. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 366–372. <https://doi.org/10.57152/malcom.v3i2.1227>
- Novianti, F., Aisyah Yasmin, Y. R., & Novitasari, D. C. R. (2022). Penerapan Algoritma Fuzzy C-Means (FCM) dalam Pengelompokan Provinsi di Indonesia berdasarkan Indikator Penyakit Menular Manusia. *JUMANJI (Jurnal Masyarakat Informatika Unjani)*, 6(1), 23. <https://doi.org/10.26874/jumanji.v6i1.103>
- Rathee, J., Kaur, P., & Singh, A. (2022). Evaluating the performance of Fuzzy Clustering using different distance metrics for Image Segmentation. *2022 International Conference for Advancement in Technology (ICONAT)*, 1–5. <https://doi.org/10.1109/ICONAT53423.2022.9725950>
- Sanusi, W., Zaky, A., & Afni, B. N. (2020). Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk. *Journal of Mathematics, Computations, and Statistics*, 2(1), 47. <https://doi.org/10.35580/jmathcos.v2i1.12458>
- Sharifnia, A. M., Kpormegbey, D. E., Thapa, D. K., & Cleary, M. (2026). A Primer of Data Cleaning in Quantitative Research: Handling Missing Values and Outliers. *Journal of Advanced Nursing*, 82(1), 970–975. <https://doi.org/10.1111/jan.16908>
- Sholeh, M., Lestari, U., & Andayati, D. (2024). Comparison of Feature Selection with Information Gain Method in Decision Tree, Regression Logistic and Random Forest Algorithms. *Journal of Applied Business and Technology*, 5(3), 146–153. <https://doi.org/10.35145/jabt.v5i3.155>
- Srirahayu, A., & Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, 14(1). <https://doi.org/10.36982/jiig.v14i1.2981>
- Surya Maulana, A., Nazir, A., Handayani, L., & Afrianty, I. (2023). Penerapan Algoritma Fuzzy C-Means untuk Melihat Pola Penerima Beasiswa Bank Indonesia. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 3(6), 670–679. <https://doi.org/10.30865/klik.v3i6.788>
- Zahrani Putri, A., Afdal, M., Monalisa, S., & Permana, I. (2023). Penerapan Algoritma Fuzzy C-Means Pada Segmentasi Pelanggan B2B dengan Model LRFM. *Jurnal Media Informatika Budidarma*, 7(3), 1423. <https://doi.org/10.30865/mib.v7i3.6150>