

Perbandingan Algoritma *K-Means* dan *Fuzzy C-Means Clustering* Pada Penilaian Kelurahan Berprestasi Kota Jambi

Yuga Pramudya¹, Jasmir², Kurniabudi³

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: ¹yugaapramudya@gmail.com, ²jasmir@unama.ac.id, ³kurniabudi@unama.ac.id

Email Penulis Korespondensi: yugaapramudya@gmail.com

Submitted :
10 April 2025

Revision :
24 Mei 2025

Accepted:
25 September 2025

Published:
30 September 2025

Abstrak—Penilaian kelurahan berprestasi tingkat Kota Jambi menghadapi kendala dalam penyusunan kategori kelompok pembinaan dan pemilihan algoritma yang tepat untuk menangani data heterogen. Untuk mengatasi permasalahan tersebut, penelitian yang akan dilakukan menerapkan teknik klusterisasi menggunakan algoritma *K-Means* dan *Fuzzy C-Means* guna menetapkan segmentasi kebutuhan pembinaan berdasarkan kedekatan nilai yang homogen. Sebanyak 204 data penilaian kelurahan dianalisis menggunakan *Python*, menghasilkan klusterisasi yang hampir setara dalam kualitas pemisahan klaster. Algoritma *K-Means* memperoleh *silhouette score* sebesar 0.3854, sedikit lebih tinggi dibandingkan *Fuzzy C-Means* yang mencapai 0.3831. Kedua algoritma membentuk tiga klaster dengan rata-rata nilai 82, 72, dan 61. Klaster 0 mendapatkan penghargaan dan promosi pengembangan karir, klaster 1 difokuskan pada diklat manajemen dan reformasi birokrasi, sementara klaster 2 memerlukan pembinaan berupa *coaching clinic* dan bimbingan teknis. Hasil penelitian menunjukkan bahwa *K-Means* lebih unggul dalam kesederhanaan dan kejelasan distribusi data. Dengan teknik klusterisasi dapat membantu Pemerintah Kota Jambi, khususnya Bagian Tata Pemerintahan Setda, dalam merancang strategi pembinaan berbasis data untuk meningkatkan kualitas kelurahan.

Kata Kunci: Klusterisasi; Kelurahan; *K-Means*; *Fuzzy C-Means*; *Python*

Abstract—The assessment of outstanding urban villages at the city level in Jambi faces challenges in defining categories for coaching groups and selecting the appropriate algorithm to handle heterogeneous data. To address this issue, clustering techniques using *K-Means* and *Fuzzy C-Means* algorithms were applied to segment coaching needs based on homogeneous value proximity. This study analyzed 204 sub-district assessment data using *Python*, producing clustering results with nearly equivalent cluster separation quality. The *K-Means* algorithm achieved a *silhouette score* of 0.3854, slightly higher than *Fuzzy C-Means* at 0.3831. Both algorithms formed consistent cluster patterns with average values of 82, 72, and 61, classified into three clusters: (1) Cluster 0 receives awards and career development promotions, (2) Cluster 1 focuses on management training and bureaucratic reform, and (3) Cluster 2 requires coaching clinics and technical guidance. The findings indicate that *K-Means* is more advantageous due to its simplicity, effectiveness in handling linear datasets, and clear data distribution. This clustering approach supports the Jambi City Government, particularly the Regional Secretariat Governance Section, in designing data-driven coaching strategies to enhance the quality of sub-district development.

Keywords: Clustering; Sub-district; *K-Means*; *Fuzzy C-Means*; *Python*

1. PENDAHULUAN

Perkembangan teknologi yang semakin pesat memungkinkan peningkatan dalam pemanfaatan data secara lebih optimal untuk mendukung proses pengambilan keputusan yang lebih akurat, efisien, dan berbasis fakta [1]. Dalam hal ini, *machine learning* sebagai pendekatan analitik berbasis algoritma membantu dalam menganalisis data kompleks [2], dengan penerapan tersebut dapat memberikan hasil informasi data termasuk dalam kebutuhan segmentasi kelompok penilaian yang berkaitan dengan penyelenggaraan kegiatan pembinaan di bidang pemerintahan daerah tingkat kecamatan dan kelurahan.

Penilaian Kelurahan Berprestasi tingkat Kota Jambi Tahun 2024 diselenggarakan oleh Bagian Tata Pemerintahan Sekretariat Daerah (Setda) Kota Jambi, dengan melibatkan tim juri yang terdiri dari perwakilan setiap Organisasi Perangkat Daerah (OPD) sesuai dengan bidang dan kriteria penilaian yang telah ditetapkan. Proses penilaian berlandaskan pada pelimpahan sebagian kewenangan dari Wali Kota dalam urusan pemerintahan dan pelayanan publik, sebagaimana diatur dalam Peraturan Pemerintah Republik Indonesia No. 17 Tahun 2018 tentang Kecamatan [3], dengan penelitian yang akan dilakukan diharapkan evaluasi terhadap kinerja kelurahan dapat dilakukan secara objektif dan komprehensif, guna mendorong peningkatan kualitas tata kelola pemerintahan di tingkat kelurahan serta pelayanan publik yang lebih baik bagi masyarakat.

Bagian Tata Pemerintahan Sekretariat Daerah Kota Jambi selaku penyelenggara menghadapi kendala dalam menetapkan standar kelompok pembinaan. Untuk itu, penggunaan teknik klusterisasi dapat membantu menetapkan segmentasi pembinaan dengan mengelompokkan data berdasarkan kemiripan [4], sehingga pengamatan lebih terarah. Serta penerapan algoritma *K-Means* dan *Fuzzy C-Means*, dimana *K-Means* sebagai *hard clustering* menghasilkan pengelompokan data berdasarkan kedekatan atau kemiripan dengan segmentasi yang jelas dan efisien dalam komputasi [5], sementara *Fuzzy C-Means* sebagai *soft clustering* memungkinkan

fleksibilitas dalam keanggotaan klaster [6], Dengan demikian, perbedaan antara kedua metode tersebut dapat memberikan perspektif yang lebih luas dalam analisis pola data, memungkinkan pemilihan pendekatan yang paling optimal berdasarkan tingkat akurasi, efisiensi, serta karakteristik data yang digunakan. Dalam konteks penilaian kelurahan berprestasi di Kota Jambi tahun 2024, pemilihan metode yang tepat menjadi krusial untuk memastikan proses evaluasi berjalan secara objektif dan transparan. Dengan memahami kelebihan dan keterbatasan masing-masing metode, pihak penyelenggara dapat menentukan strategi terbaik dalam mengelompokkan dan menilai kelurahan secara lebih adil dan akurat.

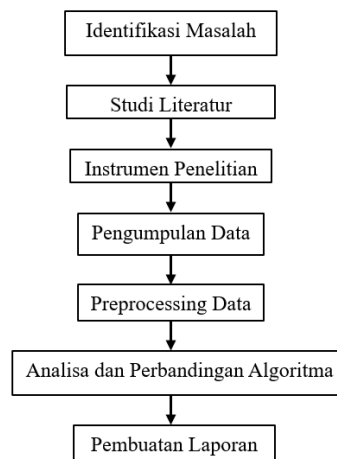
Pada tahun 2023 Ifatus Sufairoh dkk melakukan penelitian menggunakan teknik klasterisasi algoritma *K-Means*, DBSCAN, dan *Hierarchical* didapatkan hasil model terbaik yaitu *hierarchical* dengan jumlah *cluster* 2 menghasilkan *Silhouette* sebesar 0.944473 [7]. Pada tahun 2021 Fitriana Harahap dkk melakukan penelitian menggunakan teknik klasterisasi algoritma *K-Means* dan *K-Medoids* didapatkan hasil *K-Medoids Clustering*, dengan DBI (*Davis-Bouldin Index*) yang lebih rendah yaitu 0,161, sedangkan *K-Medoids* memiliki nilai DBI 0,281 [8]. Pada tahun 2022 Gideon Bartolomeus Kaligis dkk melakukan penelitian menggunakan teknik klasterisasi algoritma *K-Means*, *K-Medoids*, dan *X-Means* hasilnya *K-Means* yang unggul karena memiliki nilai *Davies-Bouldin Index* terkecil sebesar -0.377, dibandingkan *K-Medoids* (-0.930) dan *X-Means* (-0.497) [9]. Pada tahun 2023 Alvin Kurnia Rahmansyah melakukan penelitian menggunakan teknik klasterisasi algoritma *Fuzzy C-Means* dan *K-Means* hasil terbaiknya *Fuzzy C-Means* dengan *Normalization* 0.49666983222478 [10]. Tahun 2021 Anissa Enggar Pramitasari melakukan penelitian menggunakan teknik klasterisasi algoritma *Fuzzy C-Means* dan *K-Means* hasilnya *Fuzzy C-Means* lebih unggul sebesar 76% [5]. Berbeda dengan penelitian sebelumnya yang lebih berfokus pada perbandingan klasterisasi secara umum, penelitian ini lebih secara spesifik menyoroti perbedaan pendekatan *hard clustering* (*K-Means*) yang menghasilkan segmentasi tegas dan efisien [11], serta *soft clustering* (*Fuzzy C-Means*) [12] yang memungkinkan fleksibilitas keanggotaan klaster. Pendekatan dapat memberikan perspektif yang lebih luas dalam analisis pola data, memungkinkan pemilihan metode yang paling optimal sesuai dengan karakteristik data yang digunakan. Selain itu, penelitian yang dilakukan tidak hanya berfokus pada evaluasi performa algoritma secara teknis, tetapi juga menyoroti relevansi implementatifnya dalam perumusan kebijakan publik berbasis data. Dengan demikian, hasil penelitian diharapkan dapat memberikan rekomendasi yang lebih akurat dan tepat sasaran, sehingga mendukung pengambilan keputusan yang lebih efektif dalam berbagai aspek pemerintahan dan pembangunan daerah.

Dengan demikian, penelitian yang dilakukan berfokus pada tantangan dalam penerapan algoritma *K-Means* dan *Fuzzy C-Means*, khususnya dalam menilai tingkat akurasi dan efektivitas kedua metode tersebut dalam menangani variasi data yang ada. Perbandingan antara kedua algoritma diharapkan dapat memberikan wawasan lebih mendalam mengenai keunggulan serta keterbatasan masing-masing dalam proses segmentasi, sehingga dapat menentukan metode yang paling optimal untuk diterapkan dalam konteks penilaian dan pembinaan di bidang pemerintahan daerah sesuai pengelompokan berbasis *machine learning* sehingga sistem yang belajar sendiri dari data untuk membuat keputusan tanpa diprogram berulang kali [13]. Hasil yang diperoleh diharapkan dapat memberikan keluaran algoritma yang lebih optimal dalam proses segmentasi data, sehingga dapat mendukung perumusan strategi pembinaan yang lebih efektif, terarah, dan tepat sasaran.

2. METODOLOGI PENELITIAN

2.1 Kerangka Penelitian

Kerangka penelitian (*framework*) disusun dan disertai dengan langkah-langkah dalam menyelesaikan permasalahan pada penelitian. Adapun alur penelitian dapat dilihat pada gambar 1:



Gambar 1. Alur Penelitian

Berdasarkan gambar 1 dapat dijelaskan tahapan penelitian sebagai berikut:

1. **Identifikasi Masalah**
Identifikasi permasalahan dilakukan melalui analisis terhadap berita serta informasi terkait pemerintahan Kota Jambi di tingkat kelurahan. Dari hasil analisis tersebut, ditemukan berbagai tantangan dalam proses penilaian Kelurahan Berprestasi Tahun 2024, terutama dalam hal segmentasi untuk pengelompokan pembinaan yang diperlukan untuk meningkatkan efektivitas evaluasi dan pembinaan. Untuk mengatasi permasalahan ini, diterapkan teknik klusterisasi sebagai pendekatan utama, dengan penggunaan algoritma yang dirancang untuk mengelompokkan data secara lebih akurat dan sistematis. Dengan demikian, proses segmentasi dapat dilakukan secara objektif, sehingga setiap kelurahan mendapatkan pembinaan yang sesuai dengan karakteristik dan kebutuhannya [14].
2. **Studi Literatur**
Studi literatur berperan sebagai landasan utama dalam memahami penyelenggaraan pemerintahan di Kota Jambi sesuai dengan peraturan yang berlaku. Dalam proses ini, peneliti mengumpulkan dan menganalisis data dari berbagai sumber terpercaya, seperti buku, jurnal ilmiah, serta situs *web* resmi, guna membangun dasar ilmiah yang kuat. Melalui pendekatan ini, penelitian dapat menghubungkan teori dengan praktik yang diterapkan dalam pemerintahan, sehingga menghasilkan analisis yang lebih komprehensif dan relevan dengan konteks kebijakan serta implementasinya di lapangan.
3. **Instrumen Penelitian**
Beberapa instrumen penelitian yang digunakan antara lain pengamatan, wawancara, dan studi dokumen. Pengamatan dilakukan untuk memperoleh informasi langsung terkait penyelenggaraan perlombaan kelurahan berprestasi tingkat Kota Jambi tahun 2024. Wawancara mengenai pelaksanaan perlombaan dilakukan bersama Ibu Nana Widyastuti, S.STP, selaku Kasubbag Administrasi Pemerintahan pada Bagian Tata Pemerintahan Setda Kota Jambi. Studi dokumen dilakukan dengan menganalisis laporan penilaian kelurahan berprestasi yang berbentuk dataset elektronik berformat *excel*, diperoleh melalui staf Bagian Tata Pemerintahan Setda Kota Jambi.
4. **Pengumpulan Data**
Dataset yang diperoleh dalam format *excel* mencakup penilaian terhadap 68 kelurahan, yang masing-masing dinilai berdasarkan tiga bidang utama, yaitu Kasi Pem dan Pelum (Pemerintahan dan Pelayanan Umum), Kasi PMK dan Kesos (Pemberdayaan Masyarakat dan Kesejahteraan Sosial), serta Kasi Trantib (Ketentraman dan Ketertiban). Total jumlah data mencapai 204 data berbagai aspek kinerja kelurahan digunakan untuk menghitung, analisis, dan pengembangan penelitian.
5. **Preprocessing Data**
Tahapan *preprocessing* mencakup data *cleaning* dan data *selection* karena data yang diperoleh sudah dalam kondisi bersih, sehingga tidak memerlukan proses tambahan seperti data transformation atau data *integration*. Data *cleaning* melakukan pengolahan data yang tidak relevan atau tidak diperlukan disaring dan dihapus. Selain itu, data *selection* untuk memilih data yang benar-benar relevan dengan penelitian, memastikan bahwa hanya informasi atribut yang memiliki kontribusi signifikan.
6. **Analisa dan Perbandingan Algoritma**
Dalam proses klusterisasi data hasil penilaian kelurahan di Kota Jambi Tahun 2024 menggunakan dua algoritma yaitu *Fuzzy C-Means* (FCM) dan *K-Means*. Dari data mentah dikumpulkan dan diproses melalui *preprocessing* untuk memastikan kualitas data sebelum analisis. Kemudian, jumlah dari kluster tersebut ditentukan berdasarkan kategori yang yaitu baik, sedang, dan belum baik. FCM memulai proses dengan menginisialisasi nilai keanggotaan *fuzzy* untuk setiap data *point*, memungkinkan data untuk memiliki keanggotaan dengan tingkat keanggotaan yang berbeda. Selanjutnya melibatkan perhitungan keanggotaan *fuzzy* dan pembaruan pusat kluster [15]. *K-Means* dengan cara mengelompokkan data ke dalam jumlah kluster dengan meminimalkan jarak antar titik data dan pusat kluster iteratif, lalu akurasi *silhouette_score* dari *Scikit Learn Metrics* di *python* [16].
7. **Pembuatan Laporan**
Pada tahapan pembuatan laporan, kegiatan penelitian yang telah dilakukan disusun dalam bentuk laporan yang mengikuti sistematika terstruktur sesuai dengan pedoman yang berlaku. Laporan mencakup hasil analisis, proses perhitungan, serta hasil segmentasi klusterisasi dan akurasi melalui pengolahan data. Pada tahap penyusunan laporan, seluruh rangkaian kegiatan penelitian yang telah dilakukan dirangkum secara sistematis dalam bentuk laporan yang mengikuti struktur yang telah ditetapkan sesuai dengan pedoman yang berlaku. Laporan yang mencakup berbagai aspek penting, termasuk hasil analisis data, proses perhitungan yang digunakan, serta hasil segmentasi klusterisasi yang diperoleh dari penerapan algoritma yang dipilih. Selain itu, tingkat akurasi dari metode yang digunakan juga dianalisis melalui pengolahan data yang cermat, sehingga dapat memberikan gambaran yang jelas mengenai efektivitas pendekatan yang diterapkan dalam penelitian ini. Bagian utama dari laporan ini mencakup berbagai aspek penting, seperti hasil analisis data kuantitatif dan kualitatif, metode dan rumus perhitungan yang digunakan dalam proses analisis, serta hasil segmentasi dari proses klusterisasi yang diperoleh melalui penerapan algoritma tertentu.

2.2 Algoritma Klasterisasi

Klasterisasi adalah proses pengelompokan data ke dalam beberapa kelompok berdasarkan kebutuhan peneliti atau secara otomatis melalui kriteria statistik tertentu [17]. Pada penelitian yang dilakukan menggunakan algoritma *Fuzzy C-means* dan *K-Means clustering* untuk mengelompokkan data penilaian kelurahan.

1. Algoritma *Fuzzy C-Means*

Fuzzy C-Means, yang diperkenalkan oleh Jim Bezdek pada tahun 1981, adalah algoritma yang digunakan untuk mengelompokkan data berdasarkan segmentasi kluster dengan menentukan keanggotaan data dalam rentang nilai antara 0 hingga 1. Dalam proses klasterisasi yang dilakukan, data akan menjadi anggota kluster dengan derajat keanggotaan tertinggi. Untuk mengevaluasi hasil klasterisasi, digunakan nilai *F-measure*, yang melibatkan perhitungan presisi dan *recall* sebagai langkah awal dalam penilaiannya [18].

2. Algoritma *K-Means*

Algoritma *K-Means* merupakan salah satu algoritma yang dapat digunakan dalam proses penghitungan kluster, algoritma yang dikenalkan oleh J. Mac Queen tepatnya pada tahun 1967 dan melalui karya jurnal yang berjudul “*Some Methods for Classification and Analysis of Multivariate Observations*” dalam hal ini dilakukan proses pembagian terhadap sampel melalui pengujian, prosesnya secara efisien berdasarkan partisi terhadap varian kelas [19].

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data penelitian yang dipergunakan sebagai berikut dimana setiap kelurahan dinilai berdasarkan indikator yang telah ditetapkan, seperti partisipasi masyarakat (Mas), kelengkapan arsip (Ars), manajemen aset (Aset), kepegawaian (Peg), penerapan teknologi (Tech), dan dokumentasi kegiatan (Dok).

Tabel 1. Data Penelitian

No.	Kel	Mas	Ars	Aset	Peg	Tech	Dok
1	Jelmu (Kasi Pem dan Pelum)	88	90	60	73	74	83
2	Jelmu (Kasi PMK dan Kesos)	74	89	92	68	83	73
3	Jelmu (Kasi Trantib)	67	94	69	60	94	91
4	Pasir pj (Kasi Pem dan Pelum)	80	66	88	72	89	75
...
204	Bakung Jaya (Kasi Trantib)	82	62	80	70	91	89

3.1.1 Tahapan *Preprocessing*

Data *cleaning* dilakukan seperti yang ditunjukkan oleh tabel 2 untuk membersihkan dataset penilaian kelurahan dari duplikasi, nilai kosong, atau data tidak relevan agar sesuai dengan standar analisis. Dataset dalam format *excel* diperiksa, dan entri dengan informasi tidak lengkap diperbaiki atau dihapus. Data *selection* dalam hal digunakan untuk memilih atribut yang relevan dan paling representatif dalam menilai kinerja kelurahan. Tabel 3 menunjukkan data kualitatif atau identitas yang tidak digunakan dalam analisis karena algoritma *clustering* memerlukan atribut kuantitatif atau kategori yang dapat dikonversi ke numerik.

Tabel 2. Data *Cleaning*

No.	Kel	Mas	Ars	Aset	Peg	Tech	Dok
1	Jelmu (Kasi Pem dan Pelum)	88	90	60	73	74	83
2	Jelmu (Kasi PMK dan Kesos)	74	?	92	68	83	73
3	Jelmu (Kasi Trantib)	67	94	69	-	94	91
4	Pasir pj (Kasi Pem dan Pelum)	/	66	88	72	89	75
5	Pasir pj (Kasi PMK dan Kesos)	78	75	72	63	90	63
6	Pasir pj (Kasi Trantib)	82	85	71	60	64	xxx

Tabel 3. Data *Selection*

No	Nama	Gol	Pangkat	Kelurahan
1	ST. Benny Setiawan, S.IP., ME	III/c	Penata	Kenali Besar
2	Nurbasnelly, SP	IV/a	Pembina	Simpang Rimbo
3	Faisal Iskandar, A.Md	III/c	Penata	Bagan Pete

3.2 Persiapan Data

Pada tahap persiapan data, setelah melalui data *cleaning* dan data *selection* maka data dimuat dalam dataset file CSV dengan menggunakan nama *dataset_clustering_rounded_int.csv*. File berisi informasi atribut yang mencakup Partisipasi Masyarakat, Kelengkapan Arsip, Manajemen Aset, Kepegawaian, Penerapan Teknologi, dan Dokumentasi Kegiatan. Berikut adalah lima baris dataset yang dihasilkan:

Tabel 4. Persiapan Data

Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan
83	80	85	84	95	80
60	62	63	69	66	63
85	88	82	83	78	83
88	83	81	83	87	83
63	60	60	60	60	72

3.3 Proses Klusterisasi

Pada tahap implementasi, algoritma *K-Means* dan *Fuzzy C-Means* digunakan untuk mengelompokkan data ke dalam kluster berdasarkan karakteristiknya. *K-Means* dipilih karena efisiensinya dalam membagi data dengan meminimalkan variansi internal, sementara *Fuzzy C-Means* memungkinkan keanggotaan ganda dalam kluster untuk menangani data yang lebih fleksibel.

3.3.1 Klusterisasi *K-Means*

Setelah data siap, proses klusterisasi dilakukan menggunakan algoritma *K-Means* dengan jumlah kluster sebanyak tiga ($n_{clusters}=3$) dengan menggunakan data numerik digunakan sebagai *input*. Setelah model selesai dilatih, label kluster yang dihasilkan ditambahkan sebagai kolom baru yaitu "*Cluster*" ke dalam dataset numerik. Label tersebut yang akan menunjukkan kluster berdasarkan kedekatan pusat kluster (*centroids*). Untuk menampilkan koordinat *centroid* dari setiap kluster, digunakan atribut *cluster_centers_* hasil keluaran dari model *K-Means*. Pusat kluster kemudian dicetak dalam bentuk *dataframe* menggunakan *pandas*, dengan kolom yang menunjukkan atribut-atribut numerik yang digunakan dalam klusterisasi.

Tabel 5. Titik *Centroid Cluster K-Means*

Cluster	Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan
0	74.037037	74.074074	73.703704	73.641975	72.617284	75.000000
1	83.512821	82.525641	85.115385	83.628205	83.717949	82.384615
2	61.777778	61.311111	61.311111	61.688889	62.422222	61.222222

Setelah proses klusterisasi selesai dilakukan, data asli yang telah diproses kini dilengkapi dengan kolom tambahan yang menunjukkan kluster masing-masing data. Kolom yang berisi label kluster yang diperoleh dari hasil implementasi algoritma *K-Means*. Dengan adanya label kluster, setiap baris data dapat diidentifikasi sebagai bagian dari salah satu kelompok atau kluster yang telah terbentuk sehingga memudahkan untuk melakukan analisis lebih lanjut terhadap karakteristik setiap kluster.

Tabel 6. Lima Baris Pertama Hasil Kluster *K-Means*

Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan	Kluster
83	80	85	84	95	80	1
60	62	63	69	66	63	2
85	88	82	83	78	83	1
88	83	81	83	87	83	1
63	60	60	60	60	72	2

Langkah berikut bertujuan untuk mengetahui distribusi jumlah data pada masing-masing kluster yang terbentuk dari proses *K-means*.

Tabel 7. Hasil Jumlah Kluster *K-Means*

Kluster 0	81 Data
Kluster 1	78 Data
Kluster 2	45 Data

Selanjutnya perhitungan nilai rata-rata untuk setiap atribut numerik berdasarkan kluster yang terbentuk. Hasil rata-ratanya memberikan informasi penting tentang karakteristik kluster, digunakan sebagai dasar untuk interpretasi dan analisis lebih mendalam terhadap pola-pola yang terbentuk dalam data.

Tabel 8. Rata-Rata Kluster *K-Means*

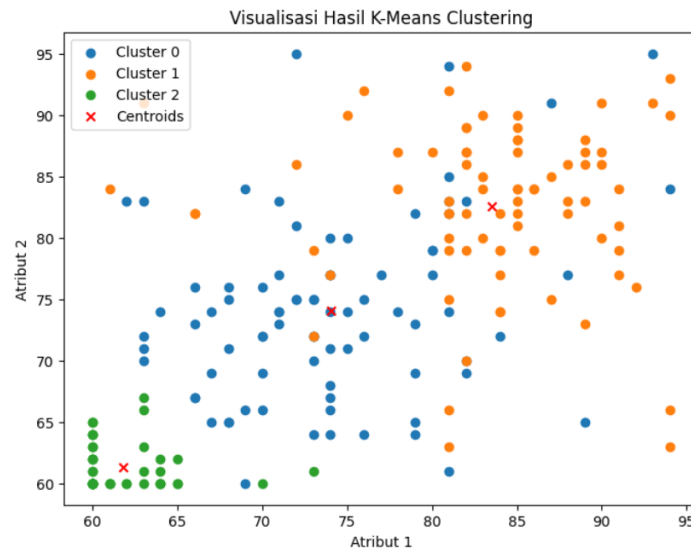
Cluster	Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan
0	74.037037	74.074074	73.703704	73.641975	72.617284	75.000000
1	83.512821	82.525641	85.115385	83.628205	83.717949	82.384615
2	61.777778	61.311111	61.311111	61.688889	62.422222	61.222222

Evaluasi dilakukan untuk menilai kualitas pengelompokan data. Salah satu metrik yang digunakan dalam evaluasi adalah *silhouette score*, yang mengukur sejauh mana data dalam satu kluster memiliki kemiripan internal yang tinggi (*cohesion*) dan berbeda secara signifikan dari data di kluster lain (*separation*).

Berikut hasil *Silhouette Score* yang dihasilkan: “0.3854036890943915”.

Nilai *Silhouette Score* berkisar antara -1 hingga 1, dengan beberapa interpretasi. Nilai yang mendekati 1 menunjukkan bahwa pengelompokan sangat baik, di mana data dalam kluster memiliki tingkat kohesi yang tinggi dan terpisah dengan jelas dari kluster lainnya. Nilai yang mendekati 0 menunjukkan bahwa data berada di perbatasan antara dua kluster, sehingga pengelompokannya kurang jelas. Sementara itu, nilai negatif menunjukkan bahwa data lebih dekat ke kluster lain dibandingkan dengan kluster anggotanya, yang mengindikasikan adanya kesalahan dalam pengelompokan.

Berikut hasil visualisasi menunjukkan distribusi kluster yang jelas dengan pusat kluster ditandai dengan simbol *x* merah. Setiap kluster diberi warna untuk membedakan data dalam satu kluster dengan kluster lainnya. Selain itu, posisi pusat kluster (*centroids*) ditampilkan menggunakan simbol “x” berwarna merah untuk memberikan gambaran mengenai lokasi pusat masing-masing kluster, diikuti dengan warna biru *cluster* 0, warna orange *cluster* 1, dan warna hijau *cluster* 2 dengan penyebaran pola kluster sudah terlihat baik dan merata, menunjukkan bahwa data telah terkelompokkan dengan jelas sesuai dengan karakteristik masing-masing kluster.



Gambar 2. Visualisasi Hasil *K-Means Clustering*

3.3.1 Klasterisasi Fuzzy *C-Means*

Setelah memuat dataset, langkah selanjutnya mempersiapkan data untuk analisis menggunakan metode *clustering fuzzy C-means*. Data mencakup kolom numerik, sehingga matriks *X_values* berisi data numerik dari dataset. Penentuan jumlah *cluster* dilakukan menetapkan *n_clusters* = 3, yang dipilih berdasarkan struktur data dan kebutuhan analisis. Proses *clustering* menggunakan metode *fuzzy C-means* menggunakan fungsi *fuzz.cmeans* dari perpustakaan *skfuzzy*. Dalam proses data dalam *X_values.T* diolah untuk mendapatkan hasil *clustering*, dengan parameter pembobotan *fuzzy* 2, toleransi kesalahan *error*=0.005, dan batas iterasi *maxiter*=1000. Hasilnya termasuk *cluster_centers*, yang menunjukkan posisi pusat dari setiap *cluster*, serta *fuzzy_membership*, yang mengindikasikan derajat keanggotaan setiap data dalam setiap *cluster*.

Selanjutnya, label *cluster* untuk setiap data ditentukan berdasarkan keanggotaan maksimum dari hasil *fuzzy_membership*. Dengan memilih *cluster* yang memiliki derajat keanggotaan tertinggi untuk setiap data, kolom *Cluster* ditambahkan ke dalam *dataframe data_numeric*, menunjukkan label *cluster* untuk masing-masing data.

Proses dapat memungkinkan interpretasi yang lebih mendalam terhadap pola-pola dalam dataset, serta memfasilitasi analisis lebih lanjut dalam penelitian. Setelah proses *clustering* selesai dilakukan menggunakan metode *fuzzy C-means*, langkah selanjutnya adalah menganalisis pusat *cluster* yang dihasilkan. Pusat *cluster* atau *cluster centers* adalah representasi dari posisi sentral setiap *cluster* dalam ruang multidimensi data. Pusat-pusat menggambarkan karakteristik umum dari data dalam setiap *cluster*, sehingga dapat digunakan untuk memahami distribusi data dan perbedaan antar *cluster*.

Tabel 9. Titik Centroid Cluster Fuzzy C-Means

Cluster	Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan
0	83.533159	82.651683	83.788267	83.585176	83.140616	82.661065
1	74.376044	74.118815	74.884479	73.688180	73.331074	75.201775
2	62.423118	62.221981	62.234895	62.493852	63.148041	61.940050

Tahap berikutnya bertujuan untuk memberikan gambaran mengenai data hasil klasterisasi, proses dilakukan dengan menampilkan lima baris pertama dari dataset dengan kolom tambahan berisi label klaster untuk memastikan label klaster berhasil diterapkan pada data untuk pemeriksaan awal pengelompokan.

Tabel 10. Lima Baris Hasil Klaster Fuzzy C-Means

Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan	Kluster
83	80	85	84	95	80	0
60	62	63	69	66	63	2
85	88	82	83	78	83	0
88	83	81	83	87	83	0
63	60	60	60	60	72	2

Analisis dapat dilakukan untuk memberikan informasi mengenai distribusi data yang dihasilkan dari setiap klaster, sehingga dapat diketahui apakah pembagian klaster telah merata atau terjadi dominasi klaster sehingga dianalisis lebih lanjut terkait karakteristik klaster. Berikut merupakan hasil yang diperoleh dari jumlah masing-masing klaster yang dihasilkan:

Tabel 11. Hasil Jumlah Masing-Masing Klaster Fuzzy C-Means

Klaster 0	77 Data
Klaster 1	81 Data
Klaster 2	46 Data

Selanjutnya menghitung nilai rata-rata (*mean*) dari setiap atribut dalam masing-masing klaster bertujuan untuk memahami karakteristik utama yang menjadi representasi setiap klaster. Dengan menghitung rata-rata fitur, dapat diidentifikasi pola umum atau ciri khas yang dimiliki oleh data dalam setiap klaster. Perhitungan rata-rata dilakukan dengan cara mengelompokkan data berdasarkan label klaster yang telah ditambahkan pada kolom *Cluster*. Setelah itu, nilai rata-rata setiap atribut dihitung menggunakan fungsi *mean()* yang diterapkan pada setiap kelompok klaster.

Tabel 12. Mean Tiap Klaster Fuzzy C-Means

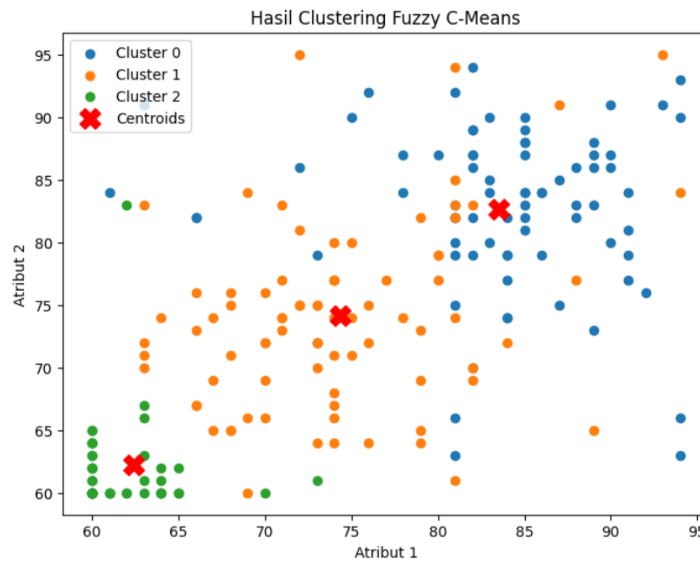
Cluster	Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan
0	83.636364	82.597403	85.051948	83.831169	83.662338	82.480519
1	74.185185	74.000000	73.975309	73.666667	72.925926	75.135802
2	61.782609	61.782609	61.456522	61.782609	62.434783	61.282609

Evaluasi dilakukan untuk menilai kualitas pengelompokan data. Salah satu metrik yang digunakan dalam evaluasi adalah *silhouette score*, yang mengukur sejauh mana data dalam satu klaster memiliki kemiripan internal yang tinggi (*cohesion*) dan berbeda secara signifikan dari data di klaster lain (*separation*).

Berikut hasil *Silhouette Score* yang dihasilkan: “0.3831”.

Setiap data dalam klaster divisualisasikan dengan warna yang berbeda, sesuai dengan label klaster yang ditentukan oleh derajat keanggotaan maksimum (*membership degree*). Posisi pusat klaster ditampilkan menggunakan simbol khusus, yaitu tanda “X” berwarna merah, diikuti dengan warna biru untuk klaster 0, warna

oranye untuk klaster 1, dan warna hijau untuk klaster 2 pola gambaran klaster sudah jelas sesuai dengan karakteristik masing-masing klaster.



Gambar 3. Visualisasi Hasil *Fuzzy C-Means Clustering*

3.3 Analisa Hasil Perbandingan Klasterisasi *K-Means* dan *Fuzzy C-Means*

Berikut merupakan hasil analisis terhadap klasterisasi menggunakan algoritma *K-Means* dan *Fuzzy C-Means* menggunakan *python*. Berikut adalah analisis komparatif yang dilakukan:

1. Distribusi Jumlah Data Klaster

Analisis terhadap hasil klasterisasi menggunakan algoritma *K-Means* dan *Fuzzy C-Means* menunjukkan perbedaan dalam distribusi data pada masing-masing klaster. Meskipun kedua algoritma menghasilkan tiga klaster dengan jumlah data yang berbeda, pola alokasi data menunjukkan hasil klasterisasi

Fuzzy C-Means	K-Means
Cluster 0: 77 data points	Jumlah Data per Cluster:
Cluster 1: 81 data points	Cluster 0: 81 data
Cluster 2: 46 data points	Cluster 1: 78 data
	Cluster 2: 45 data

Gambar 4. Hasil Distribusi Klaster

2. Rata-Rata Nilai Pada Setiap Klaster

Nilai rata-rata atribut setiap klaster menunjukkan pola yang konsisten antara kedua algoritma. *K-Means*: Klaster 1 memiliki nilai rata-rata tertinggi, diikuti klaster 0 dan klaster 2. *Fuzzy C-Means*: Klaster 0 memiliki nilai rata-rata tertinggi, diikuti klaster 1 dan klaster 2. Meski demikian, pola konsistensi dalam distribusi nilai rata-rata menunjukkan bahwa kedua metode mampu membedakan kelompok kelurahan berdasarkan karakteristik data yang dianalisis. *K-Means* maupun *Fuzzy C-Means* sama-sama efektif dalam menggambarkan segmentasi kelurahan, namun perbedaan membuat hasil interpretatif yang berbeda

Tabel 13. Perbandingan Rata-Rata Klaster *Fuzzy C-Means*

Cluster	Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan
0	83.636364	82.597403	85.051948	83.831169	83.662338	82.480519
1	74.185185	74.000000	73.975309	73.666667	72.925926	75.135802
2	61.782609	61.782609	61.456522	61.782609	62.434783	61.282609

Tabel 14. Perbandingan Rata-Rata Klaster *K-Means*

Cluster	Partisipasi Masyarakat	Kelengkapan Arsip	Manajemen Aset	Kepegawaian	Penerapan Teknologi	Dokumentasi Kegiatan
0	74.037037	74.074074	73.703704	73.641975	72.617284	75.000000
1	83.512821	82.525641	85.115385	83.628205	83.717949	82.384615
2	61.777778	61.311111	61.311111	61.688889	62.422222	61.222222

3. Evaluasi Kualitas Kluster dengan *Silhouette Score*
K-Means: 0.3854 *Fuzzy C-Means*: 0.3831. Hasilnya menunjukkan bahwa *K-Means* sedikit lebih unggul dibandingkan *Fuzzy C-Means* dalam hal kualitas pemisahan antar kluster. Namun, perbedaannya relatif kecil, sehingga dapat disimpulkan bahwa kedua algoritma memberikan hasil klusterisasi yang hampir setara dari segi kualitas pemisahan.
4. Analisa Kelebihan dan Kekurangan
K-Means: Cepat, efisien, dan cocok untuk dataset linier, tetapi rentan terhadap *outlier* dan pemilihan jumlah kluster yang tidak optimal. *Fuzzy C-Means*: Lebih fleksibel dalam menangani data kompleks, tetapi lebih lambat dan rentan terhadap *outlier* karena iterasi yang lebih banyak.
5. Interpretasi Hasil Penelitian
 Penelitian yang dilakukan mengelompokkan 68 kelurahan di Kota Jambi berdasarkan bidang-bidang seperti Kasi Pem dan Pelum, Kasi PMK dan Kesos, serta Kasi Trantib, untuk mengidentifikasi prioritas pembinaan. Hasil klusterisasi dapat dimanfaatkan oleh Bagian Tata Pemerintahan dalam merancang program pembinaan yang lebih tepat sasaran. Distribusi data yang serupa pada setiap kluster memungkinkan hasil yang digunakan sesuai kebutuhan masing-masing bidang.
 Algoritma *K-Means* menghasilkan tiga kluster dengan rincian sebagai berikut:
 - a. Kluster 0 (Sedang): 81 kelurahan (nilai rata-rata 74.04 - 75.00)
 - b. Kluster 1 (Baik): 78 kelurahan (nilai rata-rata 82.38 - 85.12)
 - c. Kluster 2 (Belum Baik): 45 kelurahan (nilai rata-rata 61.22 - 62.42)
 - d. *Silhouette Score*: 0.3854
 Distribusi kluster berdasarkan bidang:
 - a. Kasi Pem dan Pelum: Kluster 0 = 25, Kluster 1 = 26, Kluster 2 = 17
 - b. Kasi PMK dan Kesos: Kluster 0 = 31, Kluster 1 = 24, Kluster 2 = 13
 - c. Kasi Trantib: Kluster 0 = 25, Kluster 1 = 28, Kluster 2 = 15
 Sedangkan algoritma *Fuzzy C-Means* menghasilkan kluster sebagai berikut:
 - a. Kluster 0 (Baik): 77 kelurahan (nilai rata-rata 82.48 - 85.05)
 - b. Kluster 1 (Sedang): 81 kelurahan (nilai rata-rata 72.92 - 75.13)
 - c. Kluster 2 (Belum Baik): 46 kelurahan (nilai rata-rata 61.28 - 62.43)
 - d. *Silhouette Score*: 0.3831
 Distribusi kluster berdasarkan bidang:
 - a. Kasi Pem dan Pelum: Kluster 0 = 25, Kluster 1 = 25, Kluster 2 = 18
 - b. Kasi PMK dan Kesos: Kluster 0 = 24, Kluster 1 = 31, Kluster 2 = 13
 - c. Kasi Trantib: Kluster 0 = 28, Kluster 1 = 25, Kluster 2 = 15
 Pendekatan klusterisasi memberikan fleksibilitas bagi Bagian Tata Pemerintahan dalam menganalisis kinerja kelurahan dari performa terendah memerlukan prioritas pembinaan dan performa lebih baik menunjukkan kinerja yang stabil dan dapat dijadikan contoh dalam penerapan kebijakan tata kelola yang efektif. Meskipun kedua algoritma menghasilkan pola klusterisasi yang hampir sama, direkomendasikan algoritma *K-Means* sebagai metode yang lebih optimal karena memiliki akurasi yang lebih tinggi dari nilai *silhouette score* yang lebih baik. *K-Means* sederhana proses komputasi serta pemanfaatan klusterisasi diharapkan proses perencanaan, evaluasi, serta monitoring kinerja dapat lebih objektif, sistematis, dan berbasis data.

Tabel 15. Hasil Pengelompokan Kluster

Kluster 0 “Baik”	Seleksi kandidat penilaian berprestasi tingkat provinsi, promosi jabatan, diklat pengembangan karir.
Kluster 1 “Sedang”	Diklat manajemen reformasi birokrasi, diklat ketatalaksanaan, diklat manajemen kinerja, diklat kepatuhan.
Kluster 2 “Belum Baik”	<i>Coaching Clinic</i> , Bimbingan Teknis.

4. KESIMPULAN

Berdasarkan hasil penghitungan dan analisa perbandingan algoritma *K-Means* dan *Fuzzy C-Means* sebanyak 204 data penilaian kelurahan berprestasi Kota Jambi tahun 2024 menggunakan *python*, sehingga hasil analisisnya yaitu perbandingan algoritma *K-Means* dan *Fuzzy C-Means* pada 204 data penilaian kelurahan berprestasi Kota Jambi tahun 2024 menunjukkan bahwa kedua algoritma menghasilkan klusterisasi yang hampir setara dalam hal kualitas pemisahan kluster, dengan nilai *Silhouette Score K-Means* sebesar 0.3854 sedikit lebih

unggul dibandingkan *Fuzzy C-Means* sebesar 0.3831. Kedua algoritma membentuk pola kluster yang konsisten dengan rata-rata nilai 82, 72, dan 61, dimana kluster 0 diberikan penghargaan dan promosi karir, kluster 1 difokuskan pada diklat manajemen dan reformasi birokrasi, serta kluster 2 membutuhkan pembinaan lebih lanjut melalui *coaching clinic* dan bimbingan teknis. Berdasarkan hasil klusterisasi yang telah dilakukan, algoritma *K-Means* lebih unggul dari *Fuzzy C-Means*. Algoritma *K-Means* melakukan pembagian pada pusat kluster secara tetap, cocok pada dataset yang jelas dan linier, distribusi data yang konsisten, lebih sederhana & tidak terlalu kompleks dimana pusat kluster tidak berubah secara dinamis saat iterasi, sehingga mengurangi beban komputasi saat mencari setiap data ke pusat kluster baru. Pengelompokan kluster dapat membantu Pemerintah Kota Jambi melalui Bagian Tata Pemerintahan Setda dalam merancang kelompok pembinaan sesuai dengan tingkat kemampuan dan kebutuhan yang lebih berbasis data, serta strategi yang diterapkan dapat meningkatkan kualitas sasaran pada kelompok pembinaan di tingkat kelurahan.

REFERENCES

- [1] N. Shah and K. Shah, "Introduction to Data Mining," in *Practical Data Mining Techniques and Applications*, 2023. doi: 10.1201/9781003390220-1.
- [2] S. R. Durugkar, R. Raja, K. K. Nagwanshi, and S. Kumar, "Introduction to data mining," 2022. doi: 10.1002/9781119792529.ch1.
- [3] Pemerintah Republik Indonesia, *Peraturan Pemerintah Republik Indonesia Nomor 17 Tahun 2018 tentang Kecamatan*. Jakarta: Pemerintah Republik Indonesia, 2018.
- [4] R. Meng and W. He, "Data Mining," in *Flavoromics: An Integrated Approach to Flavor and Sensory Assessment*, 2023. doi: 10.1201/9781003268758-8.
- [5] A. E. Pramitasari and Y. Nataliani, "Perbandingan clustering karyawan berdasarkan nilai kinerja dengan algoritma k-means dan fuzzy c-means," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 3, 2021, doi: 10.35957/jatisi.v8i3.957.
- [6] S. Kurniawan, A. M. Siregar, and H. Y. Novita, "Penerapan algoritma k-means dan fuzzy c-means dalam mengelompokkan prestasi siswa berdasarkan nilai akademik," *Scientific Student Journal for Information, Technology and Science*, vol. IV, no. 1, 2023.
- [7] I. Sufairroh, A. C. Rani, K. Amalia, and D. Rolliawati, "Perbandingan hasil analisis clustering metode k-means, DBSCAN dan hierarchical pada data marketplace electronic phone," *JOINS (Journal of Information System)*, vol. 8, no. 1, 2023, doi: 10.33633/joins.v8i1.8016.
- [8] F. Harahap, "Perbandingan algoritma k-means dan k-medoids untuk clustering kelas siswa tunagrahita," *TIN: Terapan Informatika Nusantara*, vol. 2, no. 4, 2021.
- [9] G. B. Kaligis and S. Yulianto, "Analisa perbandingan algoritma k-means, k-medoids, dan x-means untuk mengelompokkan kinerja pegawai," *IT-Explore: Jurnal Penerapan Teknologi Informasi dan Komunikasi*, vol. 1, no. 3, 2022, doi: 10.24246/itexplore.v1i3.2022.pp179-193.
- [10] A. Kurnia, "Perbandingan algoritma k-means dan fuzzy c-means untuk clustering puskesmas berdasarkan gizi balita Surabaya," *Jurnal PROCESSOR*, vol. 18, no. 1, 2023, doi: 10.33998/processor.2023.18.1.696.
- [11] H. Prastiwi, Jeny Pricilia, and Errissya Rasywir, "Implementasi data mining untuk menentukan persediaan stok barang di mini market menggunakan metode k-means clustering," *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, vol. 2, no. 1, 2022, doi: 10.33998/jakakom.2022.2.1.34.
- [12] M. B. Johra, "Soft clustering dengan algoritma fuzzy k-means (studi kasus: pengelompokan desa di Kota Tidore Kepulauan)," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 2, 2021, doi: 10.30598/barekengvol15iss2pp385-392.
- [13] M. R. S. Alfarizi, M. Z. Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, "Penggunaan python sebagai bahasa pemrograman untuk machine learning dan deep learning," *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, vol. 2, no. 1, 2023.
- [14] K. Topuz, A. Bajaj, and I. Abdulrashid, "Interpretable machine learning," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2023. doi: 10.1201/9780367816377-16.
- [15] Sola Huddin, "Penerapan fuzzy c-means pada klusterisasi penerima bantuan pangan non tunai," *KLIK: Kajian Ilmiah Informatika dan Komputer*, Aug. 2023.
- [16] A. M. Sikana and A. W. Wijayanto, "Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering," *Jurnal Ilmu Komputer*, vol. 14, no. 2, 2021, doi: 10.24843/jik.2021.v14.i02.p01.
- [17] Amna, *Data Mining*. Padang: PT Global Eksekutif Teknologi, 2023.
- [18] A. Aziz, A. Siregar, and C. Zonyfar, "Penerapan Algoritma K-Means dan Fuzzy C-Means untuk Pengelompokan Kabupaten Kota Berdasarkan Produksi Padi di Provinsi Jawa Barat," *Scientific Student Journal for Information, Technology and Science*, vol. 3, no. 1, 2022.
- [19] E. T. Puteri, G. Kusnanto, and C. J. Thomas, "Penerapan k-means clustering untuk segmentasi pelanggan pada sistem customer relationship management di PT. Unichem Candi Indonesia," *KONVERGENSI*, vol. 15, no. 2, 2020, doi: 10.30996/konv.v15i2.3651.