

# PERBANDINGAN ALGORITMA *K-MEANS* DAN *FUZZY C-MEANS* DALAM KLASTERISASI DATA PENDIDIKAN DI WILAYAH JAKARTA

**Mohamad Ardan**

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara  
Jl. Letjen S. Parman No. 1, Tomang, Grogol Petamburan, RT.6/RW.16, Kota Jakarta Barat, Daerah Khusus  
Ibu Kota Jakarta 11440, Indonesia

*E-mail:* [mohamad.535210090@stu.untar.ac.id](mailto:mohamad.535210090@stu.untar.ac.id)

## ABSTRAK

Studi ini melihat bagaimana algoritma *K-Means* dan *Fuzzy C-Means* (FCM) berfungsi dalam klasterisasi data pendidikan di wilayah Jakarta. Penelitian bertujuan untuk menentukan teknik yang lebih efisien untuk mengelompokkan data berdasarkan atribut seperti tingkat pendidikan, jenis kelamin, dan distribusi geografis. Ini akan dilakukan dengan menganalisis pola pendidikan menggunakan kedua algoritma ini. Studi ini memeriksa kedua algoritma untuk menilai kualitas kluster menggunakan skor *Silhouette*. Hasilnya menunjukkan bahwa algoritma *K-Means* mendapatkan skor *Silhouette* sebesar 0,1549, sedangkan algoritma FCM mendapatkan skor 0,1358. Ini menunjukkan bahwa pemisahan kluster *K-Means* lebih baik. Namun, FCM lebih fleksibel dalam menangani data yang ambigu. Penelitian ini meningkatkan pengambilan keputusan berbasis data tentang alokasi dan pengembangan kebijakan.

**Kata kunci:** *K-Means*, *Fuzzy C-Means*, klasterisasi, data pendidikan, *Silhouette Score*.

## Abstract

*This study examines the performance of the K-Means and Fuzzy C-Means (FCM) algorithms in clustering educational data in Jakarta. The research aims to identify a more efficient technique for grouping data based on attributes such as education level, gender, and geographical distribution. This is achieved by analyzing educational patterns using both algorithms. The study evaluates the two algorithms to assess cluster quality using the Silhouette Score. The results indicate that the K-Means algorithm achieved a Silhouette Score of 0.1549, while the FCM algorithm scored 0.1358. This demonstrates that K-Means produces better cluster separation. However, FCM is more flexible in handling ambiguous data. This research enhances data-driven decision-making regarding resource allocation and policy development.*

**Keywords:** *K-Means*, *Fuzzy C-Means*, Clustering, Education Data, *Silhouette Score*.

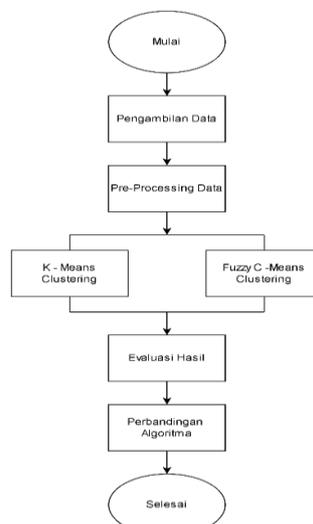
## 1. PENDAHULUAN

Pendidikan merupakan hal yang penting bagi kehidupan negara dan bangsa ini serta bagi peningkatan kualitas sumber daya manusia. Sistem pendidikan di Indonesia terdiri dari empat jenjang utama yaitu sekolah dasar (SD), sekolah menengah pertama (SMP), sekolah menengah atas (SMA), dan sekolah menengah kejuruan (SMK) [1]. Tidak semua daerah memiliki kualitas pendidikan yang sama, sehingga pemerintah harus memberikan perhatian khusus pada daerah yang dinilai berkualitas rendah. Dengan mengklasifikasikan daerah berdasarkan tingkat pendidikannya, pemerintah dapat mengidentifikasi daerah mana yang memerlukan perhatian lebih dalam penyelenggaraan pendidikan [2]. Pengelompokan ini memungkinkan pemerintah untuk menentukan daerah mana yang memiliki tingkat pendidikan lebih rendah atau lebih tinggi. Pemerintah dapat mempertimbangkan untuk memprioritaskan penyelenggaraan pendidikan di daerah-daerah tersebut sehingga daerah lain tidak tertinggal [3]. Ketidakseimbangan ini dapat berdampak pada aksesibilitas dan kualitas pendidikan masyarakat, terutama di kota-kota padat penduduk seperti Jakarta. Dalam keadaan seperti ini, sangat penting untuk melakukan analisis data pendidikan dengan teknik clustering untuk menemukan pola-pola penting yang dapat membantu keseimbangan kualitas pendidikan. Metode clustering adalah

teknik analisis data yang sering digunakan untuk menganalisis data pendidikan. Metode ini memungkinkan identifikasi pola dan pengelompokan data ke dalam kluster yang relevan, yang mempermudah interpretasi dan pengambilan keputusan [4], dengan algoritma, seperti K-means yang dikenal efisien dan dapat membagi sebuah objek menjadi kelompok berdasarkan seberapa kompatibel atau tidak kompatibelnya, dan kemudian mengelompokkan posisi objek berdasarkan nilai mean terdekatnya. Tingkat partisipasi dalam setiap kelompok dalam metode pengelompokan *fuzzy C-Means* mempengaruhi apakah suatu titik data, Fuzzy C-means juga dapat menangani data duplikat dengan lebih fleksibel [5] [6]. *Clustering* merupakan suatu teknik untuk mengelompokkan data menjadi beberapa kelompok atau cluster. Kesamaan objek dalam suatu cluster tinggi, dan kesamaan objek antar cluster berbeda rendah. Proses ini tidak dilakukan secara manual, melainkan otomatis dengan menggunakan algoritma tertentu. Objek dalam suatu cluster biasanya memiliki sifat serupa, namun objek dalam cluster lain cenderung memiliki perbedaan yang signifikan. *Clustering* sangat berguna untuk mengidentifikasi pola tersembunyi dan kelompok yang sebelumnya tidak diketahui dalam data untuk membantu analisis dan pengambilan keputusan [7]. Meskipun ada banyak algoritma klasterisasi yang tersedia untuk digunakan, setiap algoritma memiliki kelebihan dan kekurangannya sendiri. Oleh karena itu, untuk menentukan algoritma mana yang lebih baik dan menghasilkan kluster yang optimal, diperlukan perbandingan kinerja antara algoritma K-Means dan *Fuzzy C-Means* (FCM) [8]. Penelitian ini akan mengelompokkan data pendidikan dari wilayah Jakarta dengan periode, wilayah, kecamatan, kelurahan, pendidikan terakhir, dan jumlah jenis kelamin, dengan menggunakan dua algoritma K-Means dan *Fuzzy C-Means* [9]. Perbandingan antara kedua algoritma tersebut guna mendapatkan algoritma yang terbaik sesuai untuk pengelompokan data pendidikan. Diharapkan bahwa hasil penelitian ini akan memberikan wawasan yang lebih mendalam tentang keunggulan dan kelemahan masing-masing algoritma. Oleh karena itu, penelitian ini diharapkan dapat membantu menemukan pola-pola yang relevan, seperti distribusi tingkat pendidikan di wilayah Jakarta, yang membantu proses pengambilan keputusan berbasis data dalam upaya meningkatkan kualitas dan pemerataan pendidikan pada wilayah Jakarta, baik dalam skala kelurahan maupun kecamatan [10].

## 2.METODE PENELITIAN

Penelitian ini fokus pada analisis klasterisasi data menggunakan algoritma K-Means dan *Fuzzy C-Means*. Proses penelitian dibagi ke dalam beberapa tahapan utama yang sistematis untuk memastikan hasil yang valid. Tahapan tersebut meliputi pengumpulan data, pra-pemrosesan data, implementasi algoritma klasterisasi, evaluasi hasil, serta perbandingan algoritma [11]. Tahapan dapat dilihat pada Gambar 1.



Gambar 1 Proses Perbandingan Algoritma

Data yang digunakan untuk pengumpulan ini berasal dari Data Kependudukan Bersih dari Kementerian Dalam Negeri. Ini mengumpulkan informasi tentang jumlah penduduk Warga Negara Indonesia (WNI) yang direkap berdasarkan jenis pendidikan sesuai dengan kode Sistem Informasi Administrasi Kependudukan (SIAM) [12]. Situs web yang digunakan untuk pengumpulan data ini adalah Satu Data Jakarta. Dalam dataset ini termasuk variabel penting dalam analisis klusterisasi, seperti periode data, yang menjelaskan jangka waktu data yang dikumpulkan setiap satu tahun sekali dan menunjukkan perubahan data selama periode waktu tersebut. Selain itu, ada dua variabel: variabel wilayah, yang menunjukkan nama kota atau kabupaten sebagai entitas geografi yang relevan dalam analisis, dan variabel kecamatan dan kelurahan, yang masing-masing menunjukkan pembagian administratif yang lebih rendah. Tingkat pendidikan seseorang ditunjukkan oleh variabel pendidikan terakhir, yang dapat berdampak pada karakteristik sosial-ekonomi dari kluster yang terbentuk [13]. Jenis kelamin juga merupakan faktor penting untuk memahami distribusi gender dalam populasi. Selain itu, jenis kelamin seringkali menjadi faktor utama dalam studi sosial. Terakhir, angka menunjukkan jumlah total orang yang tinggal di setiap entitas geografi yang dikaji. Ini adalah ukuran utama yang menunjukkan distribusi populasi. Secara keseluruhan, semua variabel ini memberikan informasi penting untuk menentukan pola-pola tertentu dan membuat kluster-kluster yang relevan untuk penelitian sosial demografis [14]. 1-Untuk melakukan pra-pemrosesan data untuk memastikan bahwa data yang digunakan memiliki kualitas yang memadai. Tahap ini melibatkan pengecekan kelengkapan data untuk mengatasi nilai yang hilang dengan menggunakan teknik imputasi atau penghapusan data yang tidak signifikan, serta identifikasi dan penanganan outlier dengan analisis statistik. Untuk mengurangi bias pada hasil klusterisasi, data juga dinormalisasi agar memiliki skala yang sama. Selain itu, metode encoding digunakan untuk mengubah variabel kategorikal menjadi bentuk numerik [15]. Jalankan algoritma klusterisasi Dua algoritma yang digunakan adalah K-Means dan *Fuzzy C-Means*, yang digunakan pada dataset untuk membagi institusi pendidikan berdasarkan fitur penting. Sementara algoritma K-Means membagi data ke dalam kluster dengan pendekatan berbasis *centroid*, algoritma *Fuzzy C-Means* memungkinkan data menjadi anggota dari lebih dari satu kluster dalam tingkat keanggotaan tertentu. Algoritma diimplementasikan menggunakan Python, yang memiliki pustaka pendukung seperti Scikit-learn dan NumPy [16].

Pada evaluasi hasil klusterisasi. Untuk mengukur kinerja algoritma, metrik evaluasi seperti Silhouette Score untuk visualisasi hasil klusterisasi digunakan. Tujuan evaluasi ini adalah untuk menentukan algoritma mana yang paling baik untuk pada data ini. Pendekatan penelitian ini dirancang untuk memberikan hasil yang robust, mendalam, dan relevan dalam mengidentifikasi pola-pola dalam data pendidikan menggunakan metode klusterisasi berbasis algoritma.

### 2.1. Dataset

Dalam penelitian ini, Variabel penting untuk analisis klusterisasi termasuk dalam dataset ini, seperti "periode\_data", yang menjelaskan periode data yang dikumpulkan setiap satu tahun sekali dan menunjukkan perubahan data selama periode tersebut. Selain itu, ada variabel "wilayah" dan "kecamatan", masing-masing menunjukkan pembagian administratif yang lebih rendah pada tingkat yang lebih rendah, serta nama kota atau kabupaten sebagai entitas geografi yang relevan dalam analisis. Variabel "pendidikan\_terakhir" menunjukkan tingkat pendidikan terakhir seseorang, yang dapat mempengaruhi karakteristik sosial-ekonomi kelompok yang terbentuk. Jenis kelamin juga merupakan variabel penting untuk memahami distribusi gender dalam populasi; ini seringkali menjadi faktor penentu dalam studi sosial. Terakhir, jumlah menunjukkan jumlah total penduduk di setiap entitas geografi yang dianalisis. Ini adalah ukuran utama untuk menunjukkan bagaimana populasi tersebar di berbagai wilayah. Secara keseluruhan, semua variabel ini memberikan data penting untuk menentukan pola-pola tertentu dan membuat kluster-kluster yang relevan untuk penelitian sosial demografis.

Penelitian ini bertujuan untuk mengelompokkan pendidikan di wilayah Jakarta berdasarkan karakteristik signifikan menggunakan algoritma K-Means dan *Fuzzy C-Means*. Hasil dari analisis ini diharapkan mampu mengidentifikasi pola atau kelompok dalam data yang dapat dimanfaatkan untuk

mendukung pengambilan keputusan strategis, seperti alokasi sumber daya, pengembangan kebijakan, atau perencanaan program intervensi berdasarkan kebutuhan spesifik tiap klaster [16]. Dengan demikian, penelitian ini tidak hanya memberikan wawasan mendalam tentang kondisi pendidikan di Jakarta tetapi juga berkontribusi dalam merumuskan kebijakan berbasis data.

## 2.2. Pembagian Data Training dan Data Testing

Untuk memastikan bahwa model yang dibangun dapat diuji secara objektif dengan data yang belum pernah dilihat sebelumnya, pembagian data merupakan langkah penting dalam setiap proses pemodelan. Dalam penelitian ini, dua rasio pembagian data digunakan, 80:20 dan 70:30. Rasio ini membagi 80% data untuk set pelatihan model dan 20% sisanya untuk set pengujian model [17]. Ketika ingin memaksimalkan jumlah data yang digunakan untuk melatih model sambil tetap menjaga jumlah data yang cukup untuk menilai model yang representatif, rasio ini sering digunakan. Sebaliknya, pembagian 70:30 memberikan 30% data untuk pengujian dan 70% untuk pelatihan. Kedua pembagian data ini digunakan untuk mengeksplorasi perbedaan kinerja antara kedua model, karena pembagian ini lebih memperbanyak data untuk pengujian dan memberikan gambaran yang lebih kuat tentang bagaimana model akan bekerja pada data yang tidak terlihat [18]. Pembagian data yang berbeda ini juga memungkinkan untuk mengevaluasi stabilitas dan generalisasi model, serta memberikan gambaran yang lebih komprehensif tentang kemampuan model untuk bekerja pada berbagai proporsi data.

## 2.3. Perancangan Model

Penelitian ini berkonsentrasi pada penerapan dua algoritma klasterisasi yang populer K-Means dan *Fuzzy C-Means* untuk mengelompokkan data pendidikan di wilayah Jakarta. Algoritma-algoritma ini dipilih karena kemampuan mereka untuk mengelompokkan data dengan karakteristik yang beragam dan terstruktur. Tujuan pembuatan model ini adalah untuk membandingkan kinerja kedua algoritma tersebut dalam klasterisasi Data Pendidikan Wilayah Jakarta berdasarkan fitur seperti periode data, wilayah, kecamatan, kelurahan, pendidikan terakhir, jenis kelamin, jumlah penduduk.

### 2.3.1 K-Means

Prinsip dasar algoritma klasterisasi K-Means adalah membagi data menjadi sejumlah klaster berdasarkan kedekatan atau kemiripan data dalam ruang fitur [20]. Algoritma ini sering digunakan dalam berbagai jenis analisis data. Algoritma ini bekerja dengan meminimalkan variasi intra-klaster dengan mengelompokkan data ke dalam k klaster. Proses ini dimulai dengan k titik acak sebagai pusat klaster awal, dan kemudian setiap data dialokasikan ke klaster yang memiliki centroid terdekat. Setelah semua data dikelompokkan, pusat klaster dihitung ulang dengan menggunakan rata-rata data dalam klaster. Sampai posisi centroid tidak berubah secara signifikan antara iterasi, prosedur ini diulang [19]. K-Means dapat digunakan dalam analisis data pendidikan di wilayah Jakarta untuk mengidentifikasi pola pendidikan berdasarkan variabel seperti jenis kelamin, pendidikan terakhir, jumlah penduduk, dan wilayah. Penggunaan K-Means juga dapat membantu dalam pembagian populasi untuk kebijakan pendidikan yang lebih tepat sasaran. Namun demikian, ketergantungan K-Means pada jumlah klaster yang harus ditentukan sebelumnya merupakan salah satu keterbatasannya. Pemilihan jumlah klaster yang salah dapat memengaruhi kualitas hasil klasterisasi [20].

### 2.3.1 Fuzzy C-Means

*Fuzzy C-Means* (FCM) adalah algoritma klasterisasi yang mirip dengan K-Means, tetapi melihat keanggotaan data dalam klaster dengan cara yang berbeda. Berbeda dengan K-Means, yang mengalokasikan setiap data ke satu klaster tertentu, FCM membiarkan setiap data memiliki tingkat keanggotaan yang berbeda di beberapa klaster sekaligus, berdasarkan fungsi keanggotaan fuzzy. Nilai keanggotaan FCM berkisar antara 0 dan 1 [21].

FCM memiliki keunggulan dalam menemukan pola-pola yang lebih kompleks di mana data pendidikan mungkin tidak sepenuhnya termasuk dalam satu kategori yang jelas ketika digunakan

untuk data pendidikan di wilayah Jakarta [22]. Misalnya, saat menganalisis jumlah orang dengan tingkat pendidikan tertentu, beberapa wilayah atau kelompok mungkin memiliki karakteristik yang tumpang tindih. Dalam situasi seperti itu, pendekatan fuzzy dapat digunakan untuk menganalisis karakteristik ini. Salah satu keunggulan FCM adalah kemampuan untuk menangani data yang memiliki ambiguitas dalam klasifikasi dan kemampuan untuk memberikan lebih banyak informasi tentang tingkat keterlibatan data dalam setiap kluster [23]. Namun, FCM juga memiliki kelemahan, yaitu kompleksitas komputasi yang tinggi dan kebutuhan untuk memilih parameter dengan hati-hati, seperti faktor fuzziness ( $m$ ) dan jumlah kluster.

## 2 HASIL DAN PEMBAHASAN

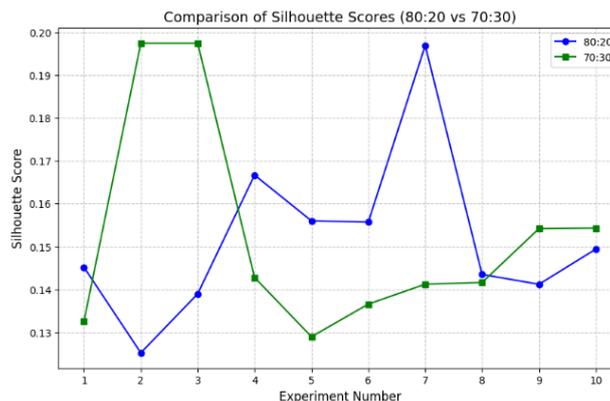
### 3.1. Hasil Evaluasi algoritma K-Means

Hasil evaluasi algoritma K-Means pada data pendidikan di wilayah Jakarta menunjukkan variasi yang cukup signifikan pada nilai Silhouette Score untuk masing-masing pembagian data. Dapat dilihat pada Tabel 1.

**Tabel 1.** Hasil Experiment K- Means Silhouette Score

Experiment ke	80:20 Silhouette Score	70:30 Silhouette Score
1	0.145176	0.132514
2	0.125186	0.19745
3	0.13896	0.19745
4	0.166675	0.142769
5	0.155994	0.128947
6	0.155712	0.13656
7	0.196932	0.141208
8	0.143483	0.141605
9	0.141189	0.154182
10	0.149391	0.154301

*Silhouette Score* digunakan untuk mengukur kualitas *cluster* yang dibuat, dengan nilai yang lebih tinggi menunjukkan partisi yang lebih baik. Untuk percobaan yang dilakukan dengan menggunakan rentang data 70: 30, nilai rata-rata skor siluet berkisar antara 0,125 hingga 0,196, dengan percobaan 7 memiliki nilai tertinggi sebesar 0,196932. Sedangkan nilai rata-rata skor siluet pada data split 80: 20 berkisar antara 0,125 hingga 0,196, dengan nilai tertinggi sebesar 0,196932 diperoleh pada Eksperimen 2.

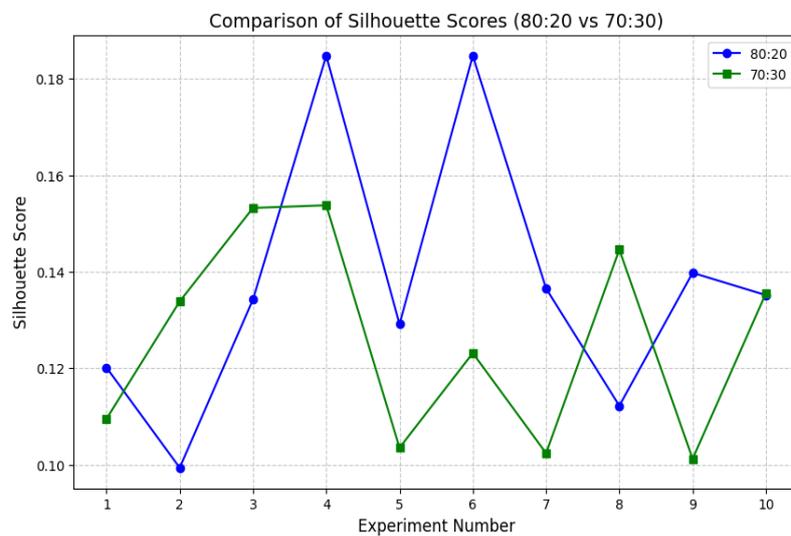


**Gambar 2** Perbandingan Hasil K-Means Silhouette Scores

Meskipun hasilnya berbeda, secara umum, pembagian data 70:30 memberikan nilai *Silhouette Score* yang lebih konsisten dan sedikit lebih tinggi dibandingkan dengan pembagian 80:20. Ini menunjukkan bahwa pembagian data dengan proporsi 70 persen untuk data latihan dan 30 persen untuk data uji cenderung memberikan kluster yang lebih baik dan terpisah dengan jelas, meskipun perbedaan antara kedua pembagian tersebut tidak terlalu mencolok. Dengan nilai *Silhouette Score* yang cukup beragam, evaluasi ini menunjukkan bahwa pembagian data dan pemilihan parameter algoritma K-Means sangat memengaruhi kualitas kluster yang dihasilkan, meskipun algoritma ini masih sederhana.

### 3.2. Hasil Evaluasi algoritma *Fuzzy C-Means*

Hasil evaluasi algoritma Fuzzy C-Means (FCM) pada data pendidikan di wilayah Jakarta menunjukkan variasi dalam nilai *Silhouette Score* yang dihasilkan pada masing-masing eksperimen, dengan pembagian data 80:20 dan 70:30, masing-masing.



**Gambar 3** Perbandingan Hasil *Fuzzy C-Means Silhouette Scores*

Pada hasil evaluasi algoritma *Fuzzy C-Means* yang digunakan untuk mengukur seberapa baik kluster terbentuk, dengan nilai yang lebih tinggi menunjukkan pemisahan kluster yang lebih besar. Pada pembagian data 80:20, nilai *Silhouette Score* 0.099385 dan 0.184755, dengan Eksperimen 4 dan 6 masing-masing menerima nilai tertinggi sebesar 0.184755. Di sisi lain, pada pembagian data 70:30, nilai *Silhouette Score* menunjukkan variasi yang lebih besar, dengan Eksperimen 3 menghasilkan nilai yang rendah.

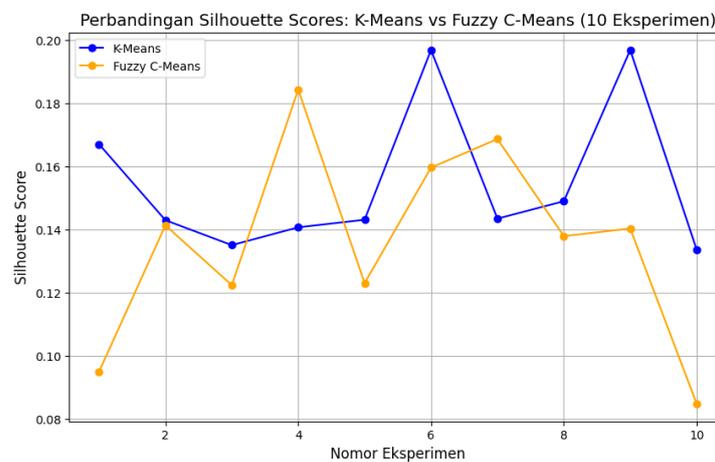
**Tabel 2** Hasil *Experiment Fuzzy C-Means Silhouette Score*

Experiment ke	80:20 Silhouette Score	70:30 Silhouette Score
1	0.120153	0.109506
2	0.099385	0.133843
3	0.134231	0.153212
4	0.184755	0.153765
5	0.129113	0.103527
6	0.184755	0.123137
7	0.136521	0.102395
8	0.11222	0.144566
9	0.139764	0.101231
10	0.135146	0.135515

Secara keseluruhan, pembagian data 70 hingga 30 memberikan nilai *Silhouette Score* yang lebih stabil dan lebih baik daripada pembagian data 80 hingga 20. Hal ini menunjukkan bahwa FCM lebih efektif dalam membentuk kluster yang lebih terpisah pada pembagian data, dengan proporsi 70 persen untuk data latih dan 30 persen untuk data uji. Namun, hasilnya mungkin berbeda dari satu eksperimen ke eksperimen lainnya, tetapi FCM mampu menghasilkan kluster dengan kualitas yang lebih baik pada sebagian besar eksperimen. Menurut evaluasi ini, algoritma Fuzzy C-Means yang memungkinkan derajat keanggotaan fuzzy pada setiap titik data mampu menangani data dengan karakteristik yang lebih kompleks dengan lebih fleksibel.

### 3.3. Perbandingan Hasil Akurasi

Hasil evaluasi menunjukkan bahwa algoritma K-Means dan *Fuzzy C-Means* memiliki kualitas kluster yang berbeda secara signifikan, dengan nilai *Silhouette Score* masing-masing 0.1549 dan 0.1358. *Silhouette Score* digunakan sebagai alat untuk mengukur seberapa baik setiap data dalam suatu kluster terpisah dari yang lain. Nilai *Silhouette Score* yang lebih tinggi menunjukkan pemisahan kluster yang lebih jelas dan lebih baik.



**Gambar 4** Perbandingan Hasil Algoritma

Pada data yang digunakan dalam penelitian ini, algoritma K-Means menunjukkan kinerja yang lebih baik dalam hal pemisahan kluster dibandingkan dengan Fuzzy C-Means, yang menunjukkan bahwa algoritma K-Means lebih baik dalam mengelompokkan data ke dalam kluster yang lebih terpisah dengan jelas. Namun, Fuzzy C-Means memiliki kelebihan dalam hal kemampuan untuk menangani ketidakpastian melalui probabilitas keanggotaan. Karakteristik masing-masing algoritma dapat menjelaskan perbedaan hasil ini. K-Means, algoritma berbasis pembagian keras, mengalokasikan setiap titik data secara tegas ke dalam salah satu kluster, yang cenderung menghasilkan pemisahan kluster yang lebih jelas dan terdefinisi. Di sisi lain, *Fuzzy C-Means* memberikan derajat keanggotaan untuk setiap titik data dalam setiap kluster, yang memungkinkan data memiliki derajat keanggotaan yang berbeda untuk lebih dari satu kluster. Meskipun metode ini memberikan fleksibilitas yang lebih besar, tumpang tindih kluster dapat meningkat, yang berdampak pada kualitas pemisahan antar kluster menurut Skor *Silhouette*.

Berdasarkan hasil ini, dapat disimpulkan bahwa, meskipun Fuzzy C-Means memiliki kelebihan dalam menangani data yang memiliki ketidakpastian atau ambiguitas, K-Means terbukti lebih efektif dalam membentuk kluster yang lebih terpisah dan lebih jelas. Oleh karena itu, mengingat hasil evaluasi yang lebih baik dalam hal pemisahan kluster, K-Means mungkin merupakan pilihan yang lebih baik untuk penelitian ini.

### 3. KESIMPULAN DAN SARAN

Tujuan dari penelitian ini adalah untuk membandingkan kinerja algoritma K-Means dan *Fuzzy C-Means* (FCM) dalam proses klasterisasi data pendidikan di wilayah Jakarta, dengan penekanan khusus pada analisis pola distribusi pendidikan terkait untuk membantu pengambilan keputusan berbasis data. Algoritma K-Means memperoleh nilai 0,1549, sedangkan algoritma *C-Means Fuzzy* memperoleh nilai 0,1358, berdasarkan evaluasi yang dilakukan menggunakan Skor Silhouette sebagai metrik utama. Hasil menunjukkan bahwa K-Means memiliki kemampuan untuk membuat klaster dengan pemisahan yang lebih jelas. Akibatnya, ini membuatnya lebih relevan untuk analisis data pendidikan yang membutuhkan pembagian kelompok yang jelas. Meskipun K-Means menunjukkan hasil yang lebih baik dalam penelitian ini, pemilihan algoritma yang paling tepat masih harus disesuaikan dengan karakteristik data dan tujuan khusus analisis.

Salah satu keunggulan utama dari pendekatan *fuzzy C-Means* adalah fleksibilitasnya dalam menangani data dengan atribut yang lebih kompleks, khususnya dalam kondisi ambiguitas, di mana satu data dapat memiliki keanggotaan di beberapa klaster pada tingkat tertentu. Metode fuzzy ini memungkinkan identifikasi pola yang lebih mendetail dan dapat berguna dalam kasus di mana kategori data tumpang tindih atau sulit didefinisikan secara kaku. Hal ini menunjukkan bahwa, Meskipun K-Means lebih baik dalam memisahkan klaster, FCM masih sangat penting untuk analisis data dengan karakteristik yang tidak terstruktur atau multikategori.

Penelitian ini sangat diharapkan untuk membantu dalam perencanaan dan pengembangan kebijakan pendidikan di Jakarta. Hasil klasterisasi yang tepat dapat digunakan oleh pemerintah untuk mengidentifikasi daerah yang membutuhkan lebih banyak perhatian dan untuk memprioritaskan alokasi sumber daya untuk meningkatkan kualitas pendidikan. Selain itu, data ini dapat digunakan untuk membuat program intervensi yang sesuai dengan kebutuhan unik setiap kelompok, seperti peningkatan akses ke pendidikan atau pengembangan infrastruktur di wilayah yang tertinggal.

Penelitian ini menghasilkan beberapa saran untuk pengembangan lebih lanjut. Pertama, penelitian selanjutnya harus menggunakan dataset yang lebih besar dan memasukkan variabel tambahan seperti status ekonomi seseorang, akses ke fasilitas pendidikan, dan tingkat partisipasi siswa. Dengan menambahkan aspek analisis ini, hasil klasterisasi dapat memberikan pemahaman yang lebih luas tentang kondisi pendidikan di wilayah Jakarta. Kedua, Anda dapat menguji algoritma klasterisasi lainnya, seperti DBSCAN atau Hierarchical Clustering, untuk melihat seberapa baik mereka menangani data pendidikan dengan berbagai pola distribusi. Terakhir, hasil klasterisasi dapat divisualisasikan dalam bentuk diagram klaster atau peta tematik untuk membuat lebih mudah dipahami bagi pengambil kebijakan. Visualisasi ini akan membuat hasil analisis lebih mudah dipahami dan diterapkan dalam kebijakan praktis. Disarankan untuk menggunakan teknik pra-pemrosesan data yang lebih canggih, seperti penanganan data tidak seimbang, deteksi outlier, dan normalisasi data yang lebih tepat, untuk meningkatkan kualitas hasil klasterisasi. Tujuan dari langkah-langkah ini adalah untuk memastikan bahwa data yang digunakan berkualitas tinggi dan representatif sehingga hasil klasterisasi lebih akurat.

Terakhir, penelitian ini diharapkan dapat berfungsi sebagai dasar untuk merumuskan kebijakan pendidikan berbasis data di Jakarta. Kebijakan ini dapat mencakup cara yang lebih efisien untuk membagi dana pendidikan, menetapkan prioritas untuk pembangunan infrastruktur pendidikan, atau mengembangkan program intervensi yang sesuai dengan kebutuhan unik setiap klaster.

## DAFTAR PUSTAKA

- [1] A. M. Rossa amelia, "data mining untuk klasterisasi provinsi di indonesia berdasarkan data pokok pendidikan nasional.," *jurnal ticom: technology of information and communication.*, vols. Volume 11, nomor 3., pp. 159-164, 2023.
- [2] Klasterisasi tingkat pendidikan di dki jakarta pada tingkat kecamatan menggunakan algoritma k-means.," *teknologia jurnal ilmiah*, vols. Vol 12, no. 4., p. 234, 2021.
- [3] P. I. Sari, "peran pendidik dalam implementasi media pembelajaran terhadap peserta didik generasi 4.0," *pros. Seminar. Nas. Pendidik. Fkip*, vols. Vol. 2, no. 1., p. 508–517, 2019.
- [4] M. A. W. S. T. Q. S. T. R. P. H. A. Jaini, "perbandingan algoritma fuzzy c-means dan k-medoids untuk pengelompokan data penjualan pada 212 mart," *i sistem informasi fakultas sains dan teknologi universitas islam negeri sultan syarif kasim riau*, 2020.
- [5] A.-k. I. H. B. Indah purnama sari, "cluster analysis using k-means algorithm and fuzzy c- means clustering for grouping students' abilities in online learning process.," *journal of computer science, information technology and telecommunication engineering (jcositte)*, vols. Vol. 2, no. 1, p. 139~144, 2021
- [6] S. S. D. Mustakim, "analisis penerimaan vaksin covid-19 berbasis fuzzy clustering machine learning di provinsi riau," *jurnal riset komputer*, vols. Vol. 8, no. 6, , p. 2407–389, 2021
- [7] P. N. P. E. T. R. Lili marlia, "analisis segmentasi pengunjung mall menggunakan algoritma k-means dan fuzzy c-means," *institut riset dan publikasi indonesia (irpi)*, pp. 240-247, 2023.
- [8] B. O. Y. Lily wulandari, "algorithm analysis of k-means and fuzzy c-means for clustering countries based on economy and health.," *faktor exacta*, vols. Vol. 15, no. 2, pp. 109-116, 2022.
- [9] M. F. F. F. J. Y. I. U. R. G. Hamdi syukron, "perbandingan k-means k-medoids dan fuzzy c-means untuk pengelompokan data pelanggan dengan model Irfm," *malcom: indonesian journal of machine learning and computer science*, vol. Vol. 2, pp. 76-83, 2022
- [10] R. N. Chairul habibi, "implementasi algoritma k-means dan c-means untuk clustering angka kemiskinan," *jurnal manajemen sistem informasi dan teknologi*, vol. Vol. 13 no. 1, p. 50, 2023
- [11] M. S. Martono, "pengklasteran dan segmentasi karakteristik donatur sedekah daring dengan teknik penambahan data.," *jurnal inovtek polbeng - seri informatika*, vols. Vol. 4, no. 1, 2020
- [12] M. S. E. L. Ika melani, "pelayanan ktp-el melalui aplikasi whatsapp di dinas kependudukan dan pencatatan sipil kabupaten fakfak," *jurnal terapan pemerintahan minangkabau*, vols. Vol. 2, no. 1., pp. 44-63, 2022
- [13] N. H. H. N. E. & P. A. Syamsiyah, "penerapan algoritma clustering dalam mengelompokkan tingkat pengangguran pada kelurahan ujung menteng jakarta timur," *jurnal komputer dan informatika*, vol. 6(2), pp. 73-80, 2022
- [14] D. Deri firmansyah, "teknik pengambilan sampel umum dalam metodologi penelitian: literature review," *jurnal ilmiah pendidikan holistik (jiph)*, vols. Vol.1, no.2, pp. 85-114, 2022.
- [15] M. A. W. Saputra, " optimasi hasil evaluasi clustering melalui kombinasi algoritma dynamic k-means dan k-means binary search centroid," (*doctoral dissertation, universitas islam negeri maulana malik ibrahim*)., 2023.
- [16] A. & S. R. Basalamah, "penerapan algoritma k-means clustering pada tingkat penyelesaian pendidikan di provinsi indonesia.," *jurnal informatika dan teknologi komputer (j-icom)*, vol. 4(2), pp. 114-121., 2023
- [17] H. F. T. R. M. I. R. L. Febrianda putra, "penerapan algoritma k-nearest neighbor menggunakan wrapper sebagai preprocessing untuk penentuan keterangan berat badan manusia," *malcom: indonesian journal of machine learning and computer science*, vol. Vol. 4, pp. 273-281, 2024
- [18] I. A. ., E. B. S. K. G. Mohd. Azhima, "penerapan metode backpropagation neural network untuk klasifikasi penyakit stroke," *klik: kajian ilmiah informatika dan komputer*, vols. Vol 4, no 6, pp. 2723-3898, 2024.
- [19] E. Akbar, "perbandingan algoritma dbscan-k means dan k means untuk pengelompokan madrasah aliyah provinsi jawa timur," (*bachelor's thesis, fakultas sains dan teknologi uin syarif hidayatullah jakarta*)., 2023
- [20] E. Akbar, "perbandingan algoritma dbscan-k means dan k means untuk pengelompokan madrasah aliyah provinsi jawa timur," *fakultas sains dan teknologi uin syarif hidayatullah jakarta*, 2023.
- [21] M. A. I. P. Syarafina dewi, "implementasi principal component analysis pada k-means untuk klasterisasi tingkat pendidikan penduduk kabupaten semaran," *jipi (jurnal ilmiah penelitian dan pembelajaran informatika*, vols. Vol. 8, no. 4, pp. 1186-1195, 2023

- [22] A. L. N. B. S. A. M. A. H. S. Firdaus, "perbandingan metode fuzzy c-means dan k-means untuk pemetaan daerah rawan kriminalitas di kota semarang," *jurnal geodesi dan geomatika*, vols. Vol. 4, no. 01, pp. 58-64, 2021
- [23] S. & M. F. Salsabila, "pola asuh orang tua single parent dalam melakukan pembinaan moral anak di desa ngoresan rt 02 rw 17 kecamatan jebres kota surakarta tahun," *doctoral dissertation, uin raden mas said surakarta*, 2022.
- [24] S. K. D. D. J. T. S. Sulartopo sulartopo, "transformasi proyek melalui keajaiban kecerdasan buatan: mengeksplorasi potensi ai dalam project management," *jurnal publikasi ilmu manajemen (jupiman)*, vols. Vol. 2, no. 2, pp. 363-392, 2023.
- [25] I. B. A. P. Made pasek agus ariawana, "prediksi nilai akhir matakuliah mahasiswa menggunakan metode k-meansclustering (studi kasus: matakuliah pemrograman dasar)," *jurnal nasional teknologi dan sistem informasi* , pp. 122-131, 2023.