
PENERAPAN FUZZY C-MEANS UNTUK PENGELOMPOKKAN TINGKAT KUALITAS PENDIDIKAN DI JAWA TIMUR

Achmad Resnu Maulana

Fakultas Sains dan Teknologi, Program Studi Matematika

Universitas Islam Negeri Sunan Ampel

Email: resnum96@gmail.com

Nurissaidah Ulinnuha

Fakultas Sains dan Teknologi, Program Studi Matematika

Universitas Islam Negeri Sunan Ampel

Email: nuris.ulinnuha@uinsby.ac.id

Dian Candra Rini Novitasari

Fakultas Sains dan Teknologi, Program Studi Matematika

Universitas Islam Negeri Sunan Ampel

Email: diancrini@uinsby.ac.id

ABSTRAK

Salah satu faktor yang menentukan apakah perkembangan manusia itu positif atau negatif adalah pendidikan. Khususnya Provinsi Jawa Timur yang memiliki angka kesenjangan mutu pendidikan antar daerah dan Lembaga. Penelitian ini bertujuan untuk untuk mengelompokkan kualitas pendidikan pada kota/kabupaten di Jawa Timur berdasarkan sejumlah parameter yaitu, angka partisipasi sekolah kasar, angka partisipasi murni, rata-rata lama sekolah, angka melek huruf. Metode clustering pada penelitian ini menggunakan algoritma Fuzzy C-Means. Didapatkan *cluster* terbaik 0.65 dengan 2 cluster. Berdasarkan kualitas pendidikan, temuan mengungkapkan bahwa Jawa Timur terbagi menjadi dua *cluster* kota/kabupaten: *cluster* 1 (dengan tingkat kualitas pendidikan tinggi) dengan 14 anggota, dan *cluster* 2 (dengan tingkat pendidikan rendah) dengan 24 anggota.

Kata kunci: clustering, Fuzzy C-Means, pendidikan

ABSTRACT

One of the factors that determines whether human development is positive or negative is education. Especially East Java Province, which has a disparity in the quality of education between regions and institutions. This study aims to classify the quality of education in cities/districts in East Java based on several parameters, namely gross enrollment rate, net enrollment rate, average length of schooling, literacy rate. The clustering method in this study uses the Fuzzy C-Means algorithm. The best cluster is 0.65 with 2 clusters. Based on the quality of education, the findings reveal that East Java is divided into two city/district clusters: cluster 1 (with a high level of educational quality) with 14 members, and cluster 2 (with a low education level) with 24 members.

Keywords: clustering, Fuzzy C-Means, education.

1. PENDAHULUAN

Kualitas pendidikan menjadi objek diskusi yang serius belakangan ini disebabkan kualitas pendidikan sangat menentukan sumber daya manusia. Tanpa adanya kualitas pendidikan yang baik, akan kecil harapan meraih indeks pembangunan yang bermutu. Selain itu kualitas pendidikan juga menjadi tolak ukur sebuah negara untuk dijadikan kategori sebagai negara maju. Upaya peningkatan kualitas pendidikan di Indonesia

perlu adanya peningkatan kualitas pendidikan di daerah-daerah terlebih dahulu. Terutama di daerah-daerah yang terpencil dan pembagian subsidi dana pendidikan yang belum merata [1].

Jawa Timur sendiri, kualitas pendidikan masih sangat timpang antara kota dan desa. Menurut Khofifah, Gubernur Jawa Timur, hanya sekitar 32.3 persen lulusan SMA yang melanjutkan ke jenjang perguruan tinggi dan dua puluh persen lulusan SMK yang melanjutkan ke pendidikan tinggi. Selain itu, terdapat kesenjangan yang tinggi mengenai kualitas pendidikan antara lembaga dan kota/kabupaten [2].

Hal tersebut diperlukan analisis pengelompokan *clustering* kualitas pendidikan di Provinsi Jawa Timur untuk mengetahui kota/kabupaten mana saja yang kualitas pendidikannya sudah maju dan masih tertinggal. Proses *clustering* kualitas pendidikan dibutuhkan sebuah metode yaitu salah Fuzzy C-Means *Clustering*. Algoritma Fuzzy C-Means adalah algoritma untuk mengelompokkan data fuzzy dan menentukan data mana yang merupakan bagian terbesar dari keanggotaannya[3]. Salah satu keunggulan dari Fuzzy C-Means adalah dapat mencapai pusat cluster yang konvergen, handal, efektif, sederhana dan dapat menganalisis data dalam jumlah besar [4][5].

Beberapa penelitian terdahulu menggunakan Fuzzy C-Means, seperti penelitian yang melakukan pengelompokan produksi daging ternak di Indonesia yang menghasilkan nilai *silhouette* 0.95 dan penelitian yang mengelompokkan produksi perkebunan di Indoensis yang menghasilkan nilai *silhouette* 0.84. Adapun dua penelitian yang membandingkan metode K-Means Clustering dengan Fuzzy C-Means, penelitian yang pertama mengelompokkan data kinerja karyawan yang menghasilkan nilai *silhouette* K-Means 0.52 dan Fuzzy C-Means 0.75. Penelitian kedua mengelompokkan data pemetaan motivasi mahasiswa yang menghasilkan nilai PCI K-Means 0.289 dan Fuzzy C-Means 0.509. Dari penyampaian di atas, peneliti bermaksud untuk membuat analisis kualitas pendidikan pada kota/kabupaten menggunakan metode Fuzzy C-Means. Penelitian ini bertujuan untuk mengelompokkan daerah-daerah di Jawa Timur mana saja yang tergolong dengan kualitas pendidikan tinggi dan kualitas pendidikan rendah.

2. METODOLOGI PENELITIAN

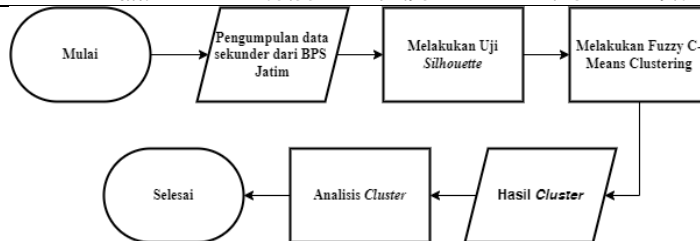
Penelitian ini memanfaatkan data sekunder dari Badan Pusat Statistik Provinsi Jawa Timur Tahun 2020. Jumlah data yang digunakan dalam penelitian ini berjumlah 38 data dan 4 variabel. Berikut ini adalah variabel penelitian yang digunakan dalam penelitian ini:

1. Angka Partisipasi Kasar (APK) usia 16 – 18 tahun, adalah Jumlah murid SMA/ sederajat dibagi dengan jumlah penduduk usia 16 – 18 tahun.
2. Angka Partisipasi Murni (APM) usia 16 – 18 tahun, adalah Jumlah murid SMA/ sederajat usia 16 – 18 tahun dibagi dengan jumlah penduduk usia 16 – 18 tahun.
3. Angka Persentase penduduk yang tamat (APT) SMA di Jawa Timur.
4. Rata – rata Lama Sekolah (RLS) usia Penududuk 15 tahun ke atas.

Data yang didapat dari BPS Jatim dapat dilihat pada Tabel 2:

Tabel 1. Kualitas Pendidikan di Jawa Timur

No	Kota/Kabupaten	APK	APM	APT	RLS
1	Pacitan	87.22	67.63	15.71	8.07
2	Ponorogo	75.95	68.15	20.37	8.01
3	Trenggalek	90.26	70.85	18.22	8.00
⋮	⋮	⋮	⋮	⋮	⋮
38	Batu	105.33	72.93	24.18	9.01



Gambar 1. Flowchart Penelitian

2.1. Fuzzy C- Mean Clustering

Metode yang dipakai pada penelitian berikut adalah Algoritma Fuzzy C-Means. Fuzzy C-Means clustering (FCM) adalah suatu metode yang ditemukan oleh Dun dan dikembangkan oleh Bezdek (1981) untuk pengelompokan data dimana derajat keanggotaan menentukan ada tidaknya setiap titik data dalam suatu *cluster* [6], [7]. Dengan menggunakan FCM, Setiap kelas atau kluster yang dibuat dengan tingkat spesifikasi dapat berisi data mulai dari 0 hingga 1 dari model pengelompokan fuzzy dan indeks fuzzy berdasarkan jarak Euclidean [8]. Fuzzy C-Means memiliki sifat sangat mudah, mudah dijalankan, memiliki kapasitas untuk mengumpulkan informasi lebih besar, lebih sehat terhadap anomali informasi[9] .

1. Dalam Fuzzy C-Means Clustering adalah membuat matriks berukuran n*m dengan:

n = banyak sampel data

m = atribut pada setiap data

x_{ij} = sampel data ke j

2. Mendefinisikan w (Rank), ξ (Error rate), $t = 1$ (iterasi awal), Maxiter (iterasi maks) dan c (jumlah cluster). $P_0 = 0$ (inisial fungsi objektif)
3. Membuat bilangan acak dimana k adalah 1, 2, ..., c dan i adalah 1, 2, ..., n menjadi anggota matriks yang partisi awalnya adalah U sesuai persamaan (1) dan (2)[10]:

$$Q_i = \sum_{k=1}^c \mu_{ik} \tag{1}$$

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \tag{2}$$

Dimana :

μ_{ik} = Bilangan acak

Q_i = Bilangan setiap kolom

4. Menghitung pusat cluster, dimana k_j adalah 1, 2, ..., m dan 1, 2, ..., n sesuai persamaan (3).

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * x_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \tag{3}$$

5. Menghitung fungsi objektif selama iterasi sebagai P_t sesuai persamaan (4).

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (x_{ij} - v_{kj})^2] (\mu_{ik})^w) \tag{4}$$

6. Menghitung perubahan partisi matriks dengan k adalah 1, 2, ..., c , m dan i adalah 1, 2, ..., n sesuai persamaan (5).

$$\mu_{ik} = \frac{[\sum_{j=1}^m (x_{ij} - V_{kj})^2]^{-1}}{\sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - V_{kj})^2]^{-1}} \tag{5}$$

2.2. Silhouette Coefficient

Silhouette Coefficient adalah metode untuk menentukan berhasil atau tidaknya pengelompokan [11][12]. Tabel yang dibuat oleh Kaufman dan Rousseeuw digunakan sebagai evaluasi nilai koefisien *silhouette* ditunjukkan pada Tabel 1 [13][14].

Tabel 2. Tabel Kaufman

<i>Nilai Silhouette Coefficient</i>	<i>Interpretasi</i>
0.71 – 1.00	Struktur yang kuat
0.51 – 0.7	Struktur yang baik/sesuai
0.26 – 0.50	Struktur yang lemah
≤ 0.25	Tidak ada struktur

Berikut adalah persamaan dari *Silhouette Coefficient* yang ditunjukkan pada persamaan (6)

$$S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \tag{6}$$

Dimana:

S_i = *Silhouette Score*

b_i = Jarak rata-rata minimum antara objek dalam cluster yang berbeda

a_i = Jarak rata-rata minimum antara objek dalam cluster yang sama

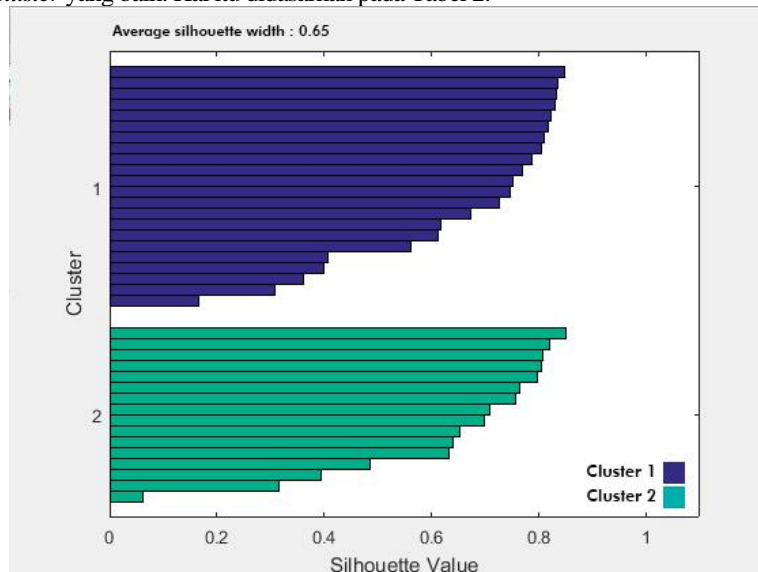
3. HASIL DAN PEMBAHASAN

3.1. Uji Silhouette

Tabel 3. Nilai Silhouette untuk setiap jumlah k cluster

	2 Cluster	3 Cluster	4 Cluster	5 Cluster
SI	0.65	0.63	0.55	0.48

Untuk pengujian jumlah k *clustering* yang optimal, perlu dilakukan pengujian nilai *silhouette* satu persatu dari 2 *cluster* sampai 5 *cluster*. Dari hasil Tabel 4 didapat *cluster* 2 sebagai *cluster* terbaik dengan hasil struktur *cluster* yang baik. Hal itu didasarkan pada Tabel 2.



Gambar 2. Silhouette Coefficient

Pada grafik *Silhouette* pada Gambar 2, nilai *silhouette* tidak terdapat nilai yang negatif. Ini menunjukkan bahwa semua data tepat masuk ke dalam *cluster*. Pada label Y mempresentasikan posisi data pada suatu *cluster*. Sedangkan pada label X mempresentasikan nilai *silhouette* setiap data.

3.2. Analisis Fuzzy C-Means Clustering

Metode Fuzzy C-Means akan digunakan untuk menghitung data, sehingga hasil *clustering* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Cluster Membership

Cluster	Anggota
1	Kab. Jombang, Sidoarjo, Kab. Mojokerto, Kota. Mojokerto, Gresik, Batu, Kota. Pasuruan, Kota. Blitar, Kota. Malang, Surabaya, Kota. Kediri, Kota. Malang, Magetan, Kota. Probolinggo, Kota. Madiun.
2	Ponorogo, Pamekasan, Kab. Kediri, Sumenep, Nganjuk, Pacitan, Trenggalek, Kab. Blitar, Sampang, Bojonegoro, Kab. Madiun, Ngawi, Kab. Malang, Pasuruan, Lumajang, Jember, Bondowoso, Situbondo, Kab. Probolinggo, Tuban, Lamongan, Bangkalan, Tulungagung, Banyuwangi.

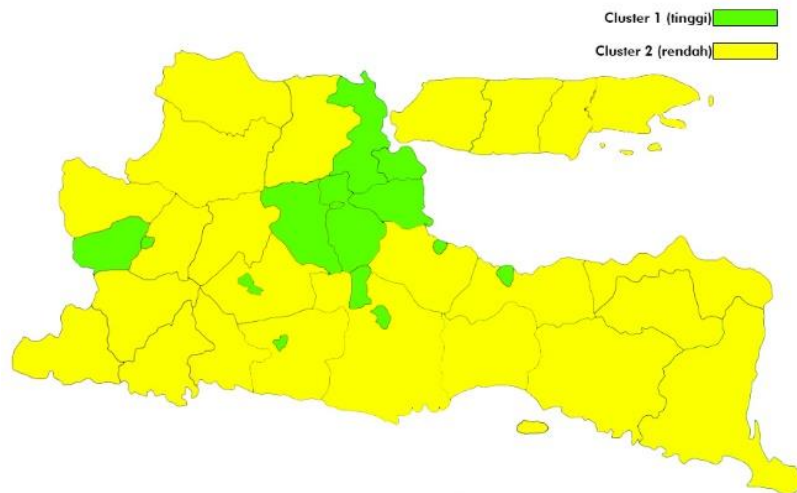
Tabel 5. Final Cluster Center

	APK	APM	APT	RLS
Cluster 1	97.84	74.20	29.67	9.47
Cluster 2	76.28	58.11	20.31	7.58

Pada Tabel 5. Untuk variabel APK terlihat bahwa *cluster 1* (97.84) > *cluster 2* (76.28). Hal tersebut berarti bahwa *cluster 1* (Kab. Jombang, Sidoarjo, Kab. Mojokerto, dll) memiliki nilai APK yang lebih tinggi dari pada *cluster 2* (Ponorogo, Pamekasan, Kab. Kediri, dll). Pada variabel APM terlihat bahwa *cluster 1* (74.20) > *cluster 2* (58.11). Hal tersebut berarti bahwa *cluster 1* (Kab. Jombang, Sidoarjo, Kab. Mojokerto, dll) memiliki nilai APM yang lebih tinggi dari pada *cluster 2* (Ponorogo, Pamekasan, Kab. Kediri, dll).

Untuk variabel APT terlihat bahwa *cluster 1* (29.67) > *cluster 2* (20.31). Hal tersebut berarti bahwa *cluster 1* (Kab. Jombang, Sidoarjo, Kab. Mojokerto, dll) memiliki nilai APT yang lebih tinggi dari pada *cluster 2* (Ponorogo, Pamekasan, Kab. Kediri, dll). Pada variabel RLS terlihat bahwa *cluster 1* (9.47) > *cluster 2* (7.58). Hal tersebut berarti bahwa *cluster 1* (Kab. Jombang, Sidoarjo, Kab. Mojokerto, dll) memiliki nilai RLS yang lebih tinggi dari pada *cluster 2* (Ponorogo, Pamekasan, Kab. Kediri, dll).

Dari penjelasan di atas, dapat ditarik kesimpulan bahwa *cluster 1* (Kab. Jombang, Sidoarjo, Kab. Mojokerto, dll) merupakan *cluster* yang memiliki tingkat kualitas pendidikan tinggi. Sedangkan *cluster 2* (Ponorogo, pamekasan, Kab. Kediri, dll) merupakan *cluster* yang memiliki tingkat kualitas rendah.



Gambar 3. Visualisasi K-Means Clustering

Dari Gambar 3 didapat interpretasi bahwa *cluster* tinggi masih berpusat pada daerah Surabaya dan sekitarnya, dan daerah perkotaan seperti Kota Malang, Kota Kediri, Kota Pasuruan, Magetan dan seterusnya. Sementara mayoritas daerah Kabupaten di Jawa timur, Pulau Madura, ujung timur Pulau Jawa dan beberapa daerah lainnya masih tergolong dalam *cluster* rendah.

4. KESIMPULAN

Dari hasil dan pembahasan diatas dapat ditarik kesimpulan bahwa Metode Fuzzy C-Means dapat digunakan dengan baik pada kota/kabupaten di Jawa Timur dengan ditunjukkannya hasil plot *silhouette* yang tidak terdapat nilai yang negatif dan struktur clusternya yang baik dengan nilai *silhouette* yaitu 0.65. Jumlah 2 cluster adalah cluster terbaik dengan *cluster* 1(kualitas pendidikan tinggi) sebanyak 36.84% dengan 14 anggota dan *cluster* 2 (kualitas pendidikan rendah) sebanyak 63.16% dengan 24 anggota.

DAFTAR PUSTAKA

- [1] Kominfo, "Dinas Kominfo Pemprov Jawtim," kominfo, surabaya, 2022.
- [2] Amaluddin, "Dinas Kominfo Pemprov Jawtim," med.com, 2022.
<https://kominfo.jatimprov.go.id/berita/ketua-dpd-ri-minta-pemerintah-buat-kebijakan-tekan-kesenjangan-pendidikan>
- [3] P. N. Agustina, "Perbandingan Algoritma K-Means Dengan Algoritma Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan," J. Resti, vol. 2, no. 3, pp. 621–626, 2018.
- [4] R. K. E. Prabowo, "Optimasi Algoritma Fuzzy Clustering dengan Menggunakan Algoritma Forest Optimatation," Inf. Syst. Dev., vol. 4, no. 1, pp. 1–6, 2019.
- [5] N. Ulinuha, "Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means," J. Ilm. Teknol., vol. 1, no. 1, pp. 8–12, 2020.
- [6] Y. N. A. Franklyn, "Pengelompokan Performa Pemain Basket Dengan Seleksi Fitur Nilai Statistik Menggunakan K-means dan Fuzzy C-means," J. Penerapan Teknol. Inf. dan Komun., vol. 1, no. 1, pp. 166–178, 2022.
- [7] Yoahannes, "Analisis Perbandingan Algoritma Fuzzy C-Means," vol. 2, no. 1, pp. 151–155, 2016.
- [8] T. Jaya, "Sistem Pemilihan Perumahan dengan Metode Kombinasi Fuzzy C-Means dan Simple Additive Weighting," J. Sist. Inf. Bisnis, vol. 1, pp. 153–158, 2018.
- [9] N. L. G. P. Suwiramayanti, "'Penerapan Metode Fuzzy C-Means Untuk Pengelompokan Data Kredit," Semin. Nas. Sist. Inf. dan Teknol. Inf., pp. 390–395.
- [10] G. W. N. S. H. Rusnedy, "Identifikasi Tingkat Pemakaian Obat Menggunakan Metode Fuzzy C-Means," J. Inf. dan Teknol., vol. 3, no. 4, pp. 196–200, 2021.
- [11] U. E. N. N. F. R. D. S. A., "Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient West Java Province Poverty Data Grouping Using the K-Means Algorithm with Silhouette Coefficient," J. Teknol. Inf. Komun., vol. 9, no. 1, pp. 29–35, 2022.

- [12] H. A. S. Paembonan, "Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat," *J. Ilm. Ilmu Tek.*, vol. 6, no. 2, p. 50.
- [13] A. S. I. D. S. A. Supriadi, "Two Step Cluster Application to Classify Villages in Kabupaten Madiun Based on Village Potential Data," *J. Stat.*, vol. 10, no. 1, pp. 12 – 16.
- [14] D. C. R. Novitasari, "Analisis Clustering pada Data Luas Grup Sunspot yang Membangkitkan Flare Soft X-Ray (SXR) Menggunakan Fuzzy C-Mean," *Snr. Penelit. Semin. Nas. Mat. Ke 11 Univ. Gajah Mada*, pp. 108–117.