

IMPLEMENTASI FUZZY C-MEANS UNTUK KLASTERISASI INDEKS PEMBANGUNAN MANUSIA KABUPATEN/KOTA DI JAWA TIMUR TAHUN 2024

Ananda Azra Razali ¹⁾, Eva Yulia Puspaningrum ¹⁾, Henni Endah Wahanani ¹⁾

1) Program Studi Informatika Universitas Pembangunan Nasional Veteran Jawa Timur

E-mail: 21081010194@student.upnjatim.ac.id, evapuspaningrum.if@upnjatim.ac.id,

henniendah@upnjatim.ac.id

Abstrak

Penelitian ini bertujuan untuk mengklaster tingkat capaian Indeks Pembangunan Manusia (IPM) di kabupaten dan kota di Provinsi Jawa Timur tahun 2024 menggunakan algoritma Fuzzy C-Means (FCM). Data yang digunakan bersumber dari Badan Pusat Statistik (BPS) dan terdiri dari empat variabel utama pembentuk IPM, yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK). Sebelum proses klasterisasi, dilakukan seleksi variabel dan normalisasi menggunakan Min-Max Scaler untuk menyeragamkan rentang nilai antarindikator. Kualitas klaster dievaluasi menggunakan Davies-Bouldin Index (DBI). Nilai DBI terendah diperoleh pada tiga klaster dengan skor 0,8162, yang menunjukkan bahwa struktur klaster tersebut merupakan yang paling optimal. Hasil klasterisasi kemudian divisualisasikan dalam bentuk tabel dan peta spasial untuk menggambarkan sebaran capaian IPM di setiap wilayah.

Kata kunci: Fuzzy C-Means, Klasterisasi, Data Mining, Davies Bouldin Index.

Pendahuluan

Kesejahteraan masyarakat merupakan kondisi yang mencerminkan tingkat kualitas hidup suatu komunitas, yang dapat diamati melalui standar kehidupan dan kemampuan masyarakat dalam memenuhi kebutuhan dasar. Salah satu indikator penting untuk menilai kesejahteraan tersebut adalah Indeks Pembangunan Manusia (IPM), yaitu ukuran komposit yang digunakan untuk menilai sejauh mana pembangunan manusia memberikan dampak terhadap peningkatan kemampuan dasar penduduk. IPM diukur melalui empat indikator utama, yakni Umur Harapan Hidup (UHH) sebagai aspek kesehatan, Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS) sebagai aspek pendidikan, serta Pengeluaran per Kapita (PPK) sebagai representasi aspek ekonomi [1].

Dalam menganalisis IPM, tantangan utama terletak pada kesulitan membedakan kategori antarwilayah karena nilai IPM kabupaten/kota di Jawa Timur umumnya memiliki rentang yang berdekatan. Oleh karena itu, diperlukan pendekatan berbasis data mining, yang berfungsi untuk mengekstraksi pola dan pengetahuan dari kumpulan data berukuran besar [2]. Salah satu teknik data mining yang sesuai untuk menganalisis kemiripan antarwilayah adalah klasterisasi (clustering), yaitu metode yang mengelompokkan data ke dalam beberapa klaster berdasarkan karakteristik yang serupa [3].

Salah satu metode clustering adalah Fuzzy C-Means (FCM) yang diperkenalkan oleh Bezdek (1981) memungkinkan setiap objek memiliki derajat keanggotaan pada beberapa klaster [4]. Pendekatan ini dinilai lebih fleksibel karena mampu menangkap ketidakpastian atau ambiguitas dalam data, terutama ketika batas antar klaster tidak tegas [5]. Salah satu kelebihan utama dari metode FCM adalah kemampuannya dalam mengelola data yang mengandung noise dan outlier dengan lebih baik dibandingkan metode klasterisasi partisi lainnya [6].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan metode FCM dalam mengelompokkan kabupaten/kota di Provinsi Jawa Timur tahun 2024 berdasarkan kesamaan indikator IPM yang mencakup dimensi pendidikan, kesehatan, dan ekonomi. Hasil klasterisasi nantinya dievaluasi menggunakan Davies-Bouldin Index untuk memperoleh pengelompokan yang akurat dan dapat menjadi dasar dalam perumusan kebijakan pembangunan manusia di tingkat daerah.

Studi Pustaka

Data mining merupakan proses analisis data berskala besar untuk menemukan pola, hubungan, dan informasi tersembunyi yang dapat membantu dalam pengambilan keputusan berbasis data [7]. Salah satu teknik utama dalam data mining adalah klusterisasi (clustering), yaitu metode pengelompokan data ke dalam beberapa kluster berdasarkan tingkat kesamaan karakteristik antar data [8]. Salah satu algoritma klusterisasi adalah Fuzzy C-Means (FCM). FCM merupakan salah satu metode klusterisasi yang digunakan untuk mengelompokkan data berdasarkan derajat keanggotaan setiap objek terhadap suatu kluster. Setiap data memiliki nilai keanggotaan antara 0 hingga 1, di mana semakin besar nilainya menunjukkan keterikatan yang lebih kuat terhadap kluster tersebut. Metode ini diperkenalkan pada tahun 1981 sebagai pengembangan dari algoritma K-Means dengan konsep fuzzy partitioning, yang memungkinkan satu data menjadi anggota lebih dari satu kluster secara bersamaan. FCM memberikan hasil pengelompokan yang lebih fleksibel dan realistis dibandingkan metode hard clustering, karena mampu menangkap ketidakpastian serta variasi yang terdapat dalam data [9].

Davies–Bouldin Index (DBI) digunakan sebagai ukuran untuk mengevaluasi kualitas hasil proses klusterisasi. Indeks ini menilai hasil pengelompokan berdasarkan dua aspek utama, yaitu kohesi dan separasi. Kohesi menggambarkan tingkat kedekatan data dalam satu kluster terhadap pusat klusternya (*centroid*), sedangkan separasi menunjukkan jarak antar *centroid* dari masing-masing kluster yang terbentuk. Nilai DBI diperoleh dari perbandingan antara kedua komponen tersebut, di mana semakin kecil nilai DBI menunjukkan kualitas klusterisasi yang semakin baik karena kluster yang terbentuk lebih kompak dan terpisah dengan jelas [10].

Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma Fuzzy C-Means (FCM) untuk mengelompokkan Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Jawa Timur tahun 2024. Data yang digunakan diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) dan mencakup variabel-variabel pembentuk IPM, yaitu angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pengeluaran per kapita. Proses penelitian disusun secara sistematis melalui beberapa tahapan, dimulai dari pengumpulan data dan prapemrosesan, dilanjutkan dengan klusterisasi menggunakan metode Fuzzy C-Means, kemudian evaluasi hasil kluster menggunakan Davies-Bouldin Index (DBI), serta visualisasi hasil kluster dalam bentuk peta yang dijelaskan sebagai berikut:

1. Pengumpulan Data: Data diperoleh dari publikasi resmi BPS berjudul Indeks Pembangunan Manusia Provinsi Jawa Timur 2024. Data mencakup indikator pembentuk IPM, yaitu angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah dan pengeluaran per kapita untuk setiap kabupaten/kota.
2. Pra-proses Data: Tahap ini meliputi seleksi variabel IPM yang relevan, pemeriksaan kelengkapan data, serta normalisasi menggunakan Min-Max Scaler agar seluruh variabel berada pada skala yang seragam sebelum dilakukan klusterisasi.
3. Klusterisasi: Metode FCM digunakan untuk membentuk sejumlah kluster berdasarkan kemiripan nilai IPM antar kabupaten/kota. Setiap wilayah memiliki derajat keanggotaan terhadap masing-masing kluster secara fleksibel sesuai karakteristik datanya.
4. Evaluasi Hasil Kluster: Kualitas pembentukan kluster diuji menggunakan DBI guna menilai tingkat pemisahan dan kedekatan antar kluster. Nilai DBI yang lebih rendah menunjukkan hasil klusterisasi yang lebih baik.
5. Visualisasi: Hasil klusterisasi akan divisualisasikan dalam bentuk peta kluster, sehingga perbedaan kelompok IPM antar kabupaten/kota di Jawa Timur dapat dilihat secara geografis dan interpretatif.

Seluruh tahapan tersebut saling berkaitan dan dilakukan secara berurutan agar proses klusterisasi berjalan optimal. Melalui rangkaian langkah ini, diharapkan hasil yang diperoleh mampu

memberikan gambaran yang akurat dan informatif mengenai pola IPM kabupaten/kota di Jawa Timur.

Hasil dan Pembahasan

Pra-proses Data

Pada tahap prapemrosesan, data dinormalisasi menggunakan metode Min-Max Scaler agar seluruh variabel pembentuk IPM berada pada skala yang sama. Normalisasi ini dilakukan untuk memastikan bahwa tidak ada variabel yang mendominasi proses klusterisasi akibat perbedaan rentang nilai. Contoh hasil normalisasi data disajikan pada Tabel 1.

Tabel 1. Contoh Hasil Normalisasi Variabel IPM Menggunakan Min-Max Scaler

Kabupaten/ Kota	UHH	HLS	RLS	PPK
Pacitan	0.527675	0.186352	0.401138	0.032072
Ponorogo	0.726937	0.472441	0.386913	0.129806
Trenggalek	0.752768	0.170604	0.403983	0.110279
Tulungagung	0.697417	0.362205	0.512091	0.220963
Blitar	0.741697	0.181102	0.396871	0.226427

Klusterisasi dan Evaluasi Hasil Klaster

Proses klusterisasi dilakukan menggunakan algoritma Fuzzy C-Means (FCM) dengan beberapa konfigurasi jumlah klaster, yaitu $c = 3, 4$ dan 5 . Parameter fuzziness yang digunakan adalah $m = 2$, dengan batas maksimum iterasi sebanyak 1000 kali serta toleransi error sebesar 10^{-5} . Kombinasi parameter tersebut diterapkan untuk memperoleh pengelompokan yang stabil dan merepresentasikan variasi IPM antar kabupaten/kota secara optimal. Setelah proses klusterisasi, evaluasi kualitas setiap percobaan jumlah klaster dilakukan menggunakan Davies-Bouldin Index (DBI). Nilai DBI digunakan untuk menentukan konfigurasi jumlah klaster terbaik berdasarkan tingkat kedekatan dalam klaster dan pemisahan antar klaster.

Tabel 2. Nilai Davies-Bouldin Index (DBI) untuk Berbagai Jumlah Klaster

Klaster	Nilai DBI
3	0.8162
4	0.9296
5	1.0335

Berdasarkan nilai DBI yang ditampilkan pada Tabel 2, konfigurasi dengan tiga klaster menghasilkan nilai DBI terendah, yaitu sebesar 0,8162. Nilai tersebut menunjukkan bahwa klaster dengan $c=3$ memiliki kualitas pemisahan dan kedekatan antar klaster yang paling baik jika dibandingkan dengan konfigurasi lainnya. Oleh karena itu, penelitian ini menggunakan tiga klaster sebagai konfigurasi akhir. Selanjutnya, pusat masing-masing klaster dari konfigurasi terbaik tersebut akan disajikan pada Tabel 3 sebagai dasar interpretasi karakteristik tiap kelompok.

Tabel 3. Pusat Klaster Berdasarkan Hasil Klusterisasi Terbaik

Klaster	UHH	HLS	RLS	PPK
1	0.61612	0.34046	0.44918	0.24379
2	0.84716	0.76835	0.83593	0.62618
3	0.17446	0.28680	0.21104	0.11047

Berdasarkan nilai pusat klaster yang disajikan pada Tabel 3, Klaster 1 menggambarkan wilayah dengan tingkat IPM sedang, ditandai oleh nilai variabel yang berada pada kisaran menengah. Klaster 2 menunjukkan IPM tinggi, dengan nilai centroid tertinggi pada seluruh variabel, seperti angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pengeluaran per kapita. Sementara itu, Klaster 3 tergolong dalam kategori IPM rendah, yang terlihat dari nilai pusat klaster

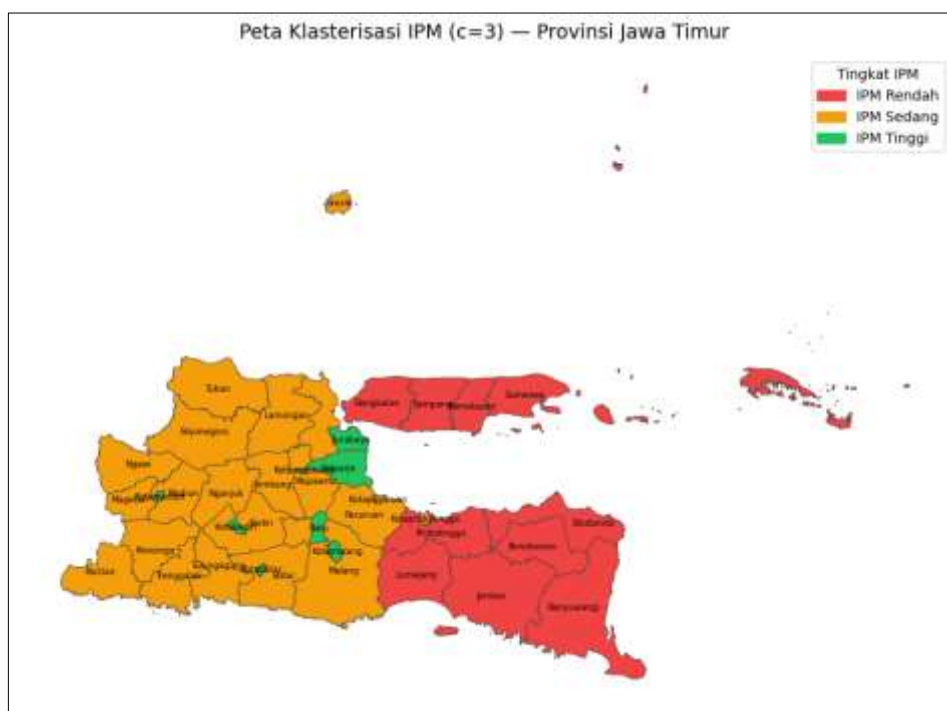
yang berada pada rentang paling kecil dibandingkan klaster lainnya. Untuk memberikan gambaran yang lebih jelas mengenai distribusi wilayah pada masing-masing klaster, hasil pengelompokannya ditampilkan pada Tabel 4, yang mencakup daftar kabupaten/kota dalam kategori IPM rendah, sedang, dan tinggi.

Tabel 4. Pusat Klaster Berdasarkan Hasil Klasterisasi Terbaik

Klaster	Kabupaten/Kota
IPM Rendah	Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Bangkalan, Sampang, Pamekasan, Sumenep, Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Kota Probolinggo, Kota Pasuruan
IPM Sedang	Sidoarjo, Kota Kediri, Kota Blitar, Kota Malang, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu

Visualisasi

Pada tahap akhir, hasil klasterisasi divisualisasikan dalam bentuk peta untuk menunjukkan pola pengelompokan IPM pada masing-masing kabupaten/kota di Jawa Timur.



Gambar 1. Peta hasil klasterisasi IPM kabupaten/kota di Jawa Timur

Pada Gambar 1 ditunjukkan distribusi klaster IPM berdasarkan hasil pengelompokan. Wilayah yang tergolong memiliki IPM rendah ditandai dengan warna merah, sedangkan wilayah dengan IPM sedang direpresentasikan oleh warna oranye. Adapun kabupaten/kota dengan tingkat IPM tinggi ditampilkan menggunakan warna hijau. Perbedaan warna ini mempermudah identifikasi kondisi pembangunan manusia di tiap daerah secara visual dan komparatif.

Kesimpulan

Berdasarkan hasil penelitian, penerapan algoritma Fuzzy C-Means (FCM) berhasil mengelompokkan Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Jawa Timur secara efektif. Evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa konfigurasi dengan tiga klaster merupakan yang paling optimal, dengan nilai DBI sebesar 0,8162. Dari hasil klasterisasi tersebut, diperoleh tiga kategori IPM, yaitu IPM rendah sebanyak 10 wilayah, IPM sedang sebanyak 20 wilayah, dan IPM tinggi sebanyak 8 wilayah. Proses visualisasi juga berjalan dengan baik, di mana peta klaster mampu menggambarkan sebaran tiap kelompok secara geografis dan memudahkan analisis spasial. Secara keseluruhan, metode yang digunakan tidak hanya menghasilkan pembagian klaster yang signifikan, tetapi juga memberikan gambaran yang jelas mengenai kondisi pembangunan manusia di Jawa Timur sebagai dasar pertimbangan kebijakan dan penelitian lanjutan. Untuk pengembangan ke depan, penelitian selanjutnya dapat mengintegrasikan variabel tambahan yang relevan, membandingkan metode FCM dengan algoritma klasterisasi lainnya, atau menerapkan teknik optimasi guna meningkatkan akurasi dan stabilitas hasil klasterisasi.

Daftar Pustaka

- [1] I. Isah, M. N. Estri, and N. Larasati, "Perbandingan Metode Fuzzy C-Means Dan Fuzzy Probabilistic C-Means Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indeks Pembangunan Manusia Tahun 2022," *Pros. Semin. Nas. Sains Data*, vol. 4, no. 1, pp. 832–841, 2024, doi: 10.33005/senada.v4i1.346.
- [2] A. Nur Khormarudin *et al.*, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 1, no. 2, pp. 116–123, 2022.
- [3] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [4] D. ARMETIYANA MARGARETTA, I. RAHMI HG, and H. YOZZA, "Pengklasteran Provinsi-Provinsi Di Indonesia Berdasarkan Indikator Yang Mempengaruhi Indeks Pembangunan Manusia Menggunakan Metode Fuzzy C-Means," *J. Mat. UNAND*, vol. 10, no. 1, pp. 79–86, 2021, doi: 10.25077/jmu.10.1.79-86.2021.
- [5] P. S. Informatika, "Penerapan Fuzzy C-Means Cluster Dalam Pengelompokan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat Pembangunan suatu negara dapat dinilai dari semakin naik atau tidaknya kesejahteraan rakyatnya . Kesejahteraan dapat diartikan sebagai suatu kead," vol. 12, no. 3, pp. 201–209, 2019.
- [6] P. F. C. D. A. N. K-means, "Perbandingan fuzzy c-means dan k-means pada klasterisasi bawang merah".
- [7] B. H. H. M Wasal Falah, "Pengintegrasian Metode Data Mining dan Prediksi untuk Analisis Fenomena Kompleks Integrating Data Mining and Prediction Methods for

- [8] N. K. Zuhail, “Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering,” *Pros. Semin. Nas. Teknol. Dan Sains*, vol. 1, pp. 200–205, 2022.
- [9] A. Yudhistira, A. A. Aldino, and D. Darwis, “Analisis Klasterisasi Penilaian Kinerja Pegawai Menggunakan Metode Fuzzy C-Means (Studi Kasus : Pengadilan Tinggi Agama bandar lampung),” *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 9, no. 1, pp. 77–82, 2022, doi: 10.21107/edutic.v9i1.17134.
- [10] M. Sholeh and K. Aeni, “Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, p. 56, 2023, doi: 10.30998/string.v8i1.16388.