Perbandingan *K-Means Clustering* dan *Fuzzy C-Means Clustering* untuk Mengelompokkan Kabupaten/Kota Berdasarkan Faktor Tingkat Pengangguran di Jawa Barat Tahun 2023

Muhammad Zaki Nabila Ibrahim *, Reny Rian Marliana

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

muhammad17zaki@gmail.com, renyrianmarliana@unisba.ac.id

Abstract. Cluster analysis groups data based on the similarity of several characteristics. The methods used in this research are K-Means and Fuzzy C-Means. K-Means has the advantage of fast clustering process and easy to implement. K-Means works by partitioning data with similar characteristics into clusters and different data into other clusters. Fuzzy C-Means (FCM) works based on the degree of membership at each data point in the cluster, so it is able to group data accurately. Unemployment is the working age group that does not yet have a job. Here are some factors that cause unemployment such as Open Unemployment Rate (TPT), Labor Force Participation Rate (TPAK), Human Development Index (HDI), and Regency / City Minimum Wage (UMK). to compare the two methods using Scatt & DensBw (S_Dbw), namely Scatt measures how far the data deviates from the cluster center, DensBw measures the density between clusters. The K-Means method is better because it has an S_Dbw value of 2.5233 smaller than FCM of 2.5297. using 2 clusters, in K-Means Cluster 1 has 9 members, namely 4 districts and 5 cities, cluster 2 has 18 members, namely 14 districts and 4 cities. while FCM cluster 1 (high unemployment rate) has 17 members, namely 14 districts and 3 cities, cluster 2 (low unemployment rate) has 10 members, namely 4 districts and 6 cities.

Keywords: *K-Means Clustering, Fuzzy C-Means (FCM).*

Abstrak. Analisis cluster mengelompokkan data berdasarkan kemiripan karakteristik. Metode yang digunakan pada penelitian ini adalah *K-Means* dan *Fuzzy C-Means* (FCM). K-Means memiliki kelebihan proses clustering cepat dan mudah diimplementasikan. *K-Means* bekerja dengan cara mempartisi data yang berkarakteristik mirip ke dalam cluster dan data berbeda ke cluster lain. FCM bekerja berdasarkan derajat keanggotaan di setiap titik data dalam cluster, sehingga mampu mengelompokkan data dengan akurat. Pengangguran adalah kelompok usia kerja yang belum memiliki pekerjaan. Berikut beberapa faktor penyebab pengangguran seperti Tingkat Pengangguran Terbuka (TPT), Tingkat Partisipasi Angkatan Kerja (TPAK), Indeks Pembangunan Manusia (IPM), dan Upah Minimum Kabupaten/Kota (UMK). untuk membandingkan kedua metode tersebut menggunakan *Scatt & DensBw* (*S_Dbw*), yaitu *Scatt* mengukur jarak penyimpangan data terhadap pusat cluster, *DensBw* mengukur kepadatan antar cluster. Metode K-Means lebih baik karena memiliki nilai *S_Dbw* sebesar 2,5233 lebih kecil daripada FCM sebesar 2,5297. menggunakan 2 cluster, pada K-Means Cluster 1 memiliki anggota sebanyak 9 anggota yaitu 4 kabupaten dan 5 kota, cluster 2 memiliki 18 anggota yaitu 14 kabupaten dan 4 kota. sedangkan FCM cluster 1 (tingkat pengangguran tinggi) memiliki 17 anggota yaitu 14 kabupaten dan 3 kota, cluster 2 (tingkat pengangguran rendah) memiliki 10 anggota yaitu 4 kabupaten dan 6 kota.

Kata Kunci: *K-Means Clustering, Fuzzy C-Means (FCM).*

*muhammad17zaki@gmail.com

A. Pendahuluan

Analisis *cluster* merupakan salah satu teknik multivariat yang bertujuan untuk mengelompokkan objek-objek dari data berdasarkan karakteristik yang dimilikinya dengan mengklasifikasikan objek yang memiliki karakteristik mendekati objek lain dalam *Cluster* yang sama [1].

K-Means Clustering bekerja dengan cara membentuk partisi data ke dalam kelompok sehingga data memiliki karakteristik sama selanjutnya dimasukan ke dalam kelompok yang sama dan data yang memiliki karakteristik berbeda dimasukan ke dalam kelompok yang lain sehingga pengelompokan data yang berbeda memiliki tingkat variasi kecil [2]. Kelebihan dari algoritma *K-Means Clustering* lebih sederhana, mudah untuk diterapkan, dan dapat digunakan pada data dengan jumlah besar. Sedangkan kekurangannya yaitu perlu menentukan jumlah *cluster* secara manual, sulit dalam mengelompokkan data dengan ukuran yang bervariasi, dan sensitif terhadap *outlier* [3].

Fuzzy C-Means Clustering merupakan salah satu teknik yang bekerja dengan cara menentukan derajat keanggotaan di setiap titik data dalam *cluster* dan dapat melakukan proses *clustering* lebih dari satu variabel sekaligus [4]. *Fuzzy C-Means Clustering* memiliki kelebihan yaitu dapat mencapai pusat *cluster* dan bersifat *unsupervised* (tidak memerlukan label), mampu mempertahankan banyak *cluster* dari data pencilan, dan mampu mengelompokkan data dengan nilai akurasi tinggi [5]. Sedangkan untuk kekurangannya dari metode *Fuzzy C-Means Clustering* yaitu sensitif terhadap pusat *cluster* awal dan mengharuskan keanggotaan kelompok dilakukan pengacakan yang menyebabkan inkonsistensi dalam hasil *clustering* [6].

K-Means Clustering merupakan metode yang paling sederhana dan cepat bekerja dengan cara meminimalkan *sum of square* suatu jarak dengan *cluster*, sedangkan *Fuzzy C-Means Clustering* merupakan salah satu metode yang bekerja dengan cara mengelompokkan suatu objek data ditentukan berdasarkan derajat keanggotaannya [7]. kedua metode tersebut sama-sama bertujuan untuk mengelompokkan objek data menjadi beberapa *cluster*. *K-Means Clustering* termasuk ke dalam algoritma *hard clustering* yang dimana setiap *cluster* nya memiliki titik pusat (pusat *cluster*),sedangkan *Fuzzy C-Means Clustering* termasuk ke dalam *soft clustering* yang artinya setiap objek data berpeluang untuk menjadi anggota lebih dari satu *cluster* [8].

Pengangguran merupakan golongan angkatan kerja yang belum memiliki kesempatan untuk memperoleh pekerjaan [9]. Berdasarkan hasil penelitian dari Ramdhan dkk (2017) menyebutkan beberapa faktor yang berpengaruh signifikan terhadap tingkat pengangguran, yaitu Upah Minimum Kota (UMK), hasil penelitian dari Mahroji & Nurkhasanah (2019) menyimpulkan bahwa UMK dan IPM berpengaruh signifikan terhadap Tingkat Pengangguran Terbuka (TPT). Penelitian dari Apriyanti dkk (2022) menunjukan bahwa Indeks Pembangunan Manusia (IPM), pertumbuhan ekonomi, investasi, dan upah minimum provinsi berpengaruh signifikan terhadap jumlah pengangguran. Sedangkan penelitian Adianita dkk (2024) menyimpulkan bahwa TPAK berpengaruh signifikan terhadap tingkat pengangguran di Indonesia.

Indeks pembangunan manusia (IPM) adalah ukuran capaian pembangunan manusia berdasarkan tiga indikator yaitu umur panjang dan sehat, pengetahuan, dan kehidupan layak. Tingkat pengangguran terbuka (TPT) merupakan persentase perbandingan antara jumlah pengangguran dengan jumlah angkatan kerja [10]. Sedangkan Partisipasi angkatan kerja (TPAK) merupakan persentase perbandingan antara angkatan kerja dengan jumlah seluruh penduduk usia kerja [11]. Upah Minimum Kabupaten/Kota (UMK) merupakan upah minimum yang ditetapkan pemerintah Kota/Kabupaten berdasarkan kondisi ekonomi dan biaya hidup di daerah tersebut [12].

Seperti yang telah disebutkan sebelumnya, penelitian ini menggunakan metode *K-Means Clustering* dan *Fuzzy C-Means Clustering* dengan indeks validitas *S_Dbw* (*Scatt & densBw*) yang bertujuan untuk mengetahui metode mana yang lebih baik dalam pengelompokkan persebaran pengangguran di Provinsi Jawa Barat tahun 2023.

Diharapkan penelitian ini dapat memberikan gambaran mengenai metode *cluster* mana yang lebih baik dan bagaimana persebaran *cluster* dari indikator TPT, TPAK, IPM, dan UMK di Provinsi Jawa Barat tahun 2023.

B. Metode

Penelitian ini menggunakan metode data kuantitatif, dan data diperoleh secara sekunder yang didapatkan melalui website Badan Pusat Statistik (BPS) Provinsi Jawa Barat dan Open Data Jabar. Menggunakan data Tingkat Pengangguran Terbuka (TPT), Tingkat Partisipasi Angkatan Kerja (TPAK), Indeks Pembangunan Manusia (IPM), dan Upah Minimum Kabupaten/Kota (UMK) tahun 2023 dengan jumlah kabupaten sebanyak 18 data dan kota sebanyak 9 data.

Normalisasi Data

Normalisasi data merupakan proses merubah nilai data menjadi nilai dengan rentang antara 0 sampai 1, yang bertujuan untuk menyetarakan data pada setiap atribut. Pada penelitian ini menggunakan metode *min-max* dilakukan dengan persamaan sebagai berikut [13]:

$$x' = \frac{x_i - min(x)}{max(x) - min(x)} \quad (1)$$

Keterangan:
x' : nilai data setelah dinormalisasi
x_i : nilai data sebelum dinormalisasi
Min(x) : nilai minimum pada data sebelum dinormalisasi
Max(x) : nilai maksimum pada data sebelum dinormalisasi

K-Means

K-Means adalah metode *cluster* yang termasuk kedalam kelompok *clustering* non-hierarki yang bekerja dengan cara mempartisi data ke dalam kelompok yang memiliki karakteristik sama dan yang berbeda akan dikelompokkan ke dalam kelompok lain. [14]. Kelebihan metode ini adalah mudah diadaptasi, proses cepat, dan mudah diimplementasikan [15]. *K* diartikan sebagai jumlah *cluster* yang akan dibuat secara acak, sedangkan *means* adalah nilai sementara dari pusat *cluster* (nilai pusat dari *cluster*), selanjutnya data diukur jaraknya menggunakan rumus *euclidean* terhadap titik pusat *cluster* sehingga dihasilkan jarak terdekat dari setiap data menggunakan pusat *cluster* (Sari dkk., 2020). Langkah-langkah *clustering* dengan menggunakan metode *K-means clustering* adalah sebagai berikut [16]:

1. Menentukan jumlah *k cluster* yang ingin dibentuk.
2. Menentukan nilai acak pada iterasi pertama yang berasal dari data yang ada sebanyak *k* untuk menghitung jarak input data terhadap masing-masing pusat *cluster* (titik pusat *cluster*) dengan menggunakan jarak *euclidean*, yaitu:

$$d(X_i, v_k) = \sqrt{\sum \left(X_{ij} - v_{jk}\right)^2} \quad (2)$$

   Keterangan:
   X_{ij} = Nilai data ke-$i = 1,2,...,$n pada atribut ke-$j = 1,2,...,m$
   v_{jk} = Pusat Cluster variabel ke-$j = 1,2,...,m$ pada Cluster ke-$k = 1,2,...,p$
3. Membuat kelompok setiap data berdasarkan jarak terdekat dengan pusat *cluster* menggunakan persamaan berikut:

$$v_{jk} = \frac{1}{n_k} \sum_{x_{ij} \in v_{jk}} x_{ij} \quad (3)$$

   Keterangan:
   v_{jk} = pusat cluster ke-$k=1,2,..,$p pada variabel ke-$j=1,2,...,m$
   x_{ij} = Nilai data atribut ke-$i=1,2,...,n,$ pada atribut ke-$j=1,2...m.$
   n_k = Banyak objek pada cluster ke-$k = 1,2,...,p$

4. Lakukan iterasi dari langkah 2 hingga 3 sampai anggota data cluster tidak ada yang berubah.

Fuzzy C-Means

Fuzzy C-Means diperkenalkan pertama kali pada tahun 1973 oleh Dunn dan di perbaiki oleh James C. Bezdek tahun 1984. Algoritma FCM banyak digunakan karena memiliki ketepatan dalam menempatkan titik pusat *cluster* tetapi memiliki kekurangan dalam nenentukan jumlah *cluster* optimal [17]. Berikut merupakan algoritma *cluster* FCM [18]:

1) Menenetukan:
   a) banyaknya *cluster* (*k*).
   b) Pembobot/pangkat (*m'*) dengan nilai umumnya berkisar antara 1,25 hingga 2 yang dimana semakin besar nilai *m'* menunjukkan bahwa pusat data *cluster* yang digunakan berada ditengah.
   c) Menentukan maksimum iterasi (*maxlter*) yaitu proses pengulangan yang berhenti ketika nilai maksimal iterasi telah tercapai.
   d) *Error* terkecil yang diharapkan (ε) merupakan batasan nilai yang membuat iterasi berhenti sesuai nilai *error* yang diharapkan.
   e) Fungsi objektif awal (P_0) merupakan suatu fungsi optimum (maksimum atau minimum), dengan nilai 1 yang berarti untuk mendapatkan nilai maksimum maupun sebaliknya.
2) Membangkitkan bilangan acak μ_{ik} sebagai elemen dari matriks partisi awal (U_0) berukuran *i x k*. Dengan μ_{ik} merupakan derajat keanggotaan objek ke-*i* pada *cluster* ke-*k*.
3) Menghitung pusat *cluster* ke-*k*, dimana menggunakan persamaan berikut:

$$v_{jk} = \frac{\sum_{i=1}^{n} \left((\mu_{ik})^{m'} x_{ij}\right)}{\sum_{i=1}^{n} (\mu_{ik})^{m'}} \quad (4)$$

Di mana:
v_{jk} = pusat *cluster* ke-*k = 1,2,...,p* pada variabel ke-*j = 1,2,...,m*
μ_{ik} : Nilai derajat keanggotaan objek ke-*i = 1,2,...,n* pada *cluster* ke-*k = 1,2,...,p*
x_{ij} : Objek data ke-*i = 1,2,...,n* pada variabel ke-*j = 1,2,...,m*
m' : pembobot/pangkat

4) Menghitung nilai fungsi objektif pada iterasi ke-*t*, *Pt* dengan menggunakan persamaan sebagai berikut:

$$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} \left(x_{ij} - v_{jk}\right)^2 \right] (\mu_{ik})^{m'} \right) \quad (5)$$

Di mana:
Pt : fungsi objektif pada iterasi ke-*t*

5) Menghitung perubahan matriks keanggotaan dengan menggunakan persamaan sebagai berikut:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m} \left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{m'-1}}}{\sum_{k=1}^{c} \left[\sum_{j=1}^{m} \left(x_{ij} - v_{kj}\right)^2\right]^{\frac{-1}{m'-1}}} \quad (6)$$

6) Ulangi langkah 3) sampai dengan langkah 5) hingga atau t > maxlter terpenuhi.

Validitas *cluster*

Validitas *cluster* bertujuan untuk mengetahui seberapa baik hasil dari pengelompokkan. indeks validitas yang digunakan adalah *S_Dbw (Scatt & densbw) Index* yang dapat memberikan hasil yang tepat dalam mengidentifikasi *cluster* [19].

Scatt menunjukkan rata-rata sebaran dalam *cluster k*, *Scatt* bertujuan untuk mengukur seberapa jauh penyimpangan data terhadap pusat *cluster*. Didefinisikan sebagai berikut:

$$Scatt(k) = \frac{1}{k}\sum_{k=1}^{p}\frac{|\sigma(v_k^j)|}{|\sigma(S)|} \quad (7)$$

Keterangan:
Scatt = rata-rata sebaran *cluster*
k = ukuran cluster
$\sigma(v_{kj})$ = varians data cluster ke-$k = 1,2,...,p$ pada variabel ke-$j = 1,2,...,m$
$\sigma(S)$ = varians dari data semua *cluster*

DensBw merupakan proses untuk mengevaluasi kepadatan (densitas) rata-rata pada wilayah antar *cluster*. didefinisikan sebagai berikut [20]:

$$Dens_{bw}(k) = \frac{1}{k(k-1)}\sum_{k=1}^{c}\frac{density(u_{ab})}{max\{density(v_a), density(v_b)\}} \quad (8)$$

Keterangan:
Densbw = kepadatan antar *cluster*
u_{ab} = titik tengah atau garis pemisah antara cluster *a* dan *b*
$max\{density(v_a),density(v_b)\}$ = nilai maksimum dari densitas v_a atau v_b.
v_a = Pusat cluster k_a
v_b = Pusat cluster k_b.

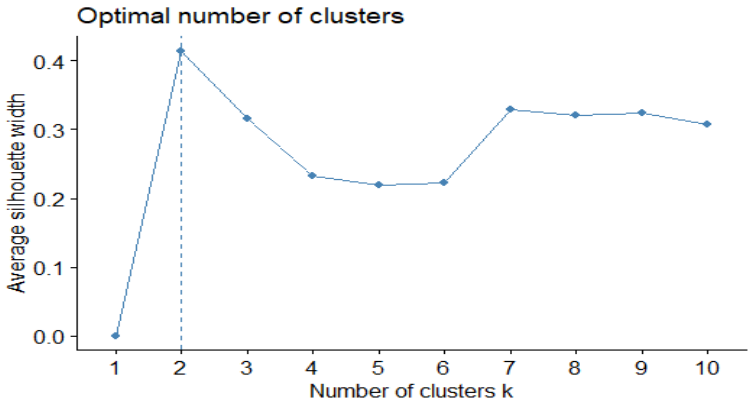
C. Hasil Penelitian dan Pembahasan

Hasil normaslisasi ditampilkan pada tabel 1 sebagai berikut:

Tabel 1. Hasil Normalisasi Data

No.	Wilayah Jawa Barat	TPT	TPAK	IPM	UMK
1	Bogor	0,7722	0,1247	0,3203	0,7936
2	Sukabumi	0,6444	0,3187	0,1013	0,426
...	...	...	...	...	...
26	Kota Tasikmalaya	0,5589	0,1918	0,4825	0,1684
27	Kota Banjar	0,4344	0,3016	0,415	0

Menentukan jumlah *Cluster* ke-*k* bertujuan untuk menentukan jumlah *Cluster* yang paling optimal, salah satu yang digunakan pada penelitian ini menggunakan metode *Silhouette* dengan Rstudio dengan *syntax* fviz_nbclust(). Hasilnya ditampilkan sebagai berikut:



Gambar 1. Visualisasi Banyaknya Cluster Optimal

Berdasarkan gambar 1, terlihat bahwa nilai rata-rata *Silhouette* tertinggi yaitu sebesar 0,4137 yang menunjukkan nilai yang optimal karena *Cohessian* (hubungan antar objek data dalam *cluster*) dan *Sparation* (pemisahan data antar *cluster*) yang tinggi sehingga dengan sebanyak 2 *cluster* data lebih representatif dibandingkan dengan banyak *cluster* lainnya.

Lakukan analisis *K-Means* dengan langkah-langkah berikut:

1. Menentukan pusat *cluster* sebanyak k menggunakan persamaan (2).
2. Mengelompokkan data berdasarkan jarak terdekat menggunakan persamaan (1).
3. Melakukan iterasi pada poin a dan b hingga pusat *cluster* tidak berubah sehingga ringkasan hasilnya sebagai berikut:

Tabel 2. Pusat *Cluster K-Means*

Cluster	**TPT**	**TPAK**	**IPM**	**UMK**
1	0,7894	0,1774	0,6555	0,8171
2	0,5496	0,3248	0,2611	0,2045

Berdasarkan besarnya pusat *Cluster* pada tabel 2 maka *Cluster* 1 masuk ke dalam kategori tingkat pengangguran tinggi, sedangkan *Cluster* 2 termasuk ke dalam kategori tingkat pengangguran rendah.

Tabel 3. Anggota *cluster K-Means*

Wilayah Jawa Barat	**TPT**	**TPAK**	**IPM**	**UMK**	**Cluster**
Bogor	0,7722	0,1247	0,3203	0,7936	1
Sukabumi	0,6444	0,3187	0,1013	0,4260	2
⋮	⋮	⋮	⋮	⋮	⋮
Kota Tasikmalaya	0,5589	0,1918	0,4825	0,1684	2
Kota Banjar	0,4344	0,3016	0,4150	0,0000	2

Dari tabel tersebut didapatkan jumlah anggota *Cluster* pada iterasi akhir $n_1 = 9$ dan $n_2 = 18$ yang artinya pada *cluster 1* terdapat 9 anggota yang terdiri dari 4 Kabupaten dan 5 Kota, sedangkan *cluster 2* terdapat 18 anggota yang terdiri dari 14 Kabupaten dan 4 Kota dengan maksimum iterasi yang dicapai sebanyak 2 kali iterasi. Kota Cimahi memiliki persentase pengangguran tertinggi, sedangkan Kabupaten Pangandaran memiliki persentase pengangguran terendah.

Lakukan analisis *Fuzzy C-Means* dengan langkah-langkah berikut:

a. Tentukan:
- Jumlah *cluster* sebanyak 2
- Bobot/ Pangkat m' = 2
- Fungsi objektif awal $P_0 = 0$
- Maksimum iterasi sebanyak 100
- *Error* terkecil yang diharapkan (ε) sebesar 10^{-5}

b. Membangkitkan bilangan random (μik).
c. Menghitung pusat *cluster* ke-*k* menggunakan persamaan (4).
d. Menghitung nilai fungsi objektif pada iterasi ke-*t*, dengan menggunakan persamaan (5).
e. Menghitung perubahan matriks keanggotaan dengan menggunakan persamaan (6).
f. Ulangi langkah c) sampai dengan langkah e) hingga $(|P_t - P_{t-1}| < \varepsilon)$
Atau *t* > *maxlter*. Sehingga hasilnya sebagai berikut:

Tabel 4. Pusat *Cluster Fuzzy C-Means*

	TPT	**TPAK**	**IPM**	**UMK**
***Cluster* 1**	0,7668	0,1847	0,6224	0,7609
***Cluster* 2**	0,5390	0,3260	0,2440	0,1929

Berdasarkan besarnya pusat *Cluster* pada tabel 4 maka *Cluster* 1 masuk ke dalam kategori tingkat pengangguran tinggi, sedangkan *Cluster* 2 termasuk ke dalam kategori tingkat pengangguran rendah.

Tabel 5. Anggota *Cluster Fuzzy C-Means*

Wilayah Jawa Barat	**TPT**	**TPAK**	**IPM**	**UMK**	***Cluster***
Bogor	0,7722	0,1247	0,3203	0,7936	1
Sukabumi	0,6444	0,3187	0,1013	0,426	2
⋮	⋮	⋮	⋮	⋮	⋮
Kota Tasikmalaya	0,5589	0,1918	0,4825	0,1684	2
Kota Banjar	0,4344	0,3016	0,415	0	2

Berdasarkan tabel 5 didapatkan jumlah anggota *Cluster* pada iterasi akhir $n_1 = 10$ dan $n_2 = 17$ yang yang artinya pada *cluster 1* terdapat 10 anggota yang terdiri dari 4 Kabupaten dan 6 Kota, sedangkan *cluster 2* terdapat 17 anggota yang terdiri dari 14 Kabupaten dan 3 Kota dengan maksimum iterasi yang dicapai sebanyak 14 kali iterasi.

Untuk mengetahui seberapa baik *cluster* yang dihasilkan dari metode *K-Means* dan *Fuzzy C-Means* serta metode yang terbaik diantara keduanya diperlukan uji validitas menggunakan metode *S_Dbw Index.* Hasil uji validitas terlihat pada tabel berikut:

Tabel 6. Hasil Perbandingan *K-Means* dan FCM Berdasarkan *S_Dbw*

S_Dbw (Scatt & DensBw)	
K-Means	FCM
2,5233	2.5297

Berdasarkan tabel 6 nilai *Scatt & DensBw* (*S_Dbw*) dinyatakan baik jika nilai yang diperoleh minimum. Pada penelitian ini menunjukkan bahwa metode *K-Means clustering* lebih baik daripada metode *Fuzzy C-Means Clustering* (FCM). Hal tersebut berdasarkan nilai *Scatt & DensBw* yang lebih kecil atau minimum.

D. Kesimpulan

Pada penelitian ini membahas mengenai pengelompokkan Kabupaten/Kota di Provinsi Jawa Barat berdasarkan faktor tingkat pengangguran tahun 2023. Berdasarkan hasil pembahasan dapat disimpulkan bahwa:

1. Penerapan *Clustering* menggunakan *K-Means Clustering* menggunakan indeks validitas *Scatt & DensBw* (*S_Dbw*) menghasilkan nilai sebesar 2,5233. Anggota *Cluster* 1 atau kelompok tingkat pengangguran tinggi terdiri dari 9 anggota *Cluster*. *Cluster* 2 atau kelompok tingkat pengangguran rendah terdiri dari 18 anggota *Cluster*.
2. Penerapan *Clustering* menggunakan *Fuzzy C-Means Clustering* menggunakan indeks validitas *Scatt & DensBw* (*S_Dbw*) menghasilkan nilai sebesar 2,5297. Anggota *Cluster* 1 atau kelompok tingkat pengangguran tinggi terdiri dari 10 anggota *Cluster*. *Cluster* 2 atau kelompok tingkat pengangguran rendah terdiri dari 17 anggota *Cluster*.
3. Kedua metode yang digunakan yaitu *K-Means Clustering* dan *Fuzzy C-Means Clustering* menghasilkan jumlah *Cluster* yang berbeda dan berdasarkan kriteria nilai minimum indeks validitas *Scatt & DensBw* (*S_Dbw*) maka pada penelitian ini metode *K-Means Clustering* dengan nilai sebesar 2,5233 lebih baik dibandingkan dengan *Fuzzy C-Means Clustering* (FCM) dengan nilai sebesar 2,5297.

Daftar Pustaka

Herliani, Fitria Dewi, and Abdul Kudus. 2023. "Penanganan Data Missing Dengan Algoritma Multivariate Imputation By Chained Equations (MICE)." *DataMath: Journal of Statistics and Mathematics* 1(1):35–42. doi:10.29313/datamath.v1i1.25.

Luhung Mustika Budiharti, and Siti Sunendiari. 2021. "Pemodelan Dan Pemetaan Jumlah Penderita Kusta Di Jawa Barat Dengan Regresi Binomial Negatif Dan Flexibly Shaped Spatial Scan Statistic." *Jurnal Riset Statistika* 1(2):99–106. doi:10.29313/jrs.v1i2.409.

P. Sihombing, Pemetaan Masalah Pembangunan Berkelanjutan dan Pertumbuhan Ekonomi Inklusif di Indonesia: Implementasi Analisis Kluster. 2018.

A. Sulistiyawati dan E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," JTK, vol. 15, no. 2, hlm. 25, Agu 2021, doi: 10.33365/jtk.v15i2.1162.

A. M. Sari, "Algoritma K- Means Clustering: Pengertian, Fungsi dan Cara Kerja," FIKTI. Diakses: 25 Februari 2024. [Daring]. Tersedia pada: https://fikti.umsu.ac.id/algoritma-k-means-clustering-pengertian-fungsi-dan-cara-kerja/

C. L. Simbolon, N. Kusumastuti, dan B. Irawan, "Clustering Lulusan Mahasiswa Matematika Fmipa Untan Pontianak Menggunakan Algoritma Fuzzy C-Means," 2013.

R. Kurniawan dan E. Prabowo, "Optimasi Algoritma Fuzzy Clustering dengan Menggunakan Algoritma Forest Optimization," Journal Information System Development (ISD), vol. 4, no. 1, Art. no. 1, Jan 2019, Diakses: 25 Februari 2024. [Daring]. Tersedia pada: https://ejournal-medan.uph.edu/isd/article/view/214

B. N. Haqiqi dan R. Kurniawan, “Analisis Perbandingan Metode Fuzzy C-Means Dan Subtractive Fuzzy C-Means,” MEDIA STATISTIKA, vol. 8, no. 2, hlm. 59–67, Des 2015, doi: 10.14710/medstat.8.2.59-67.

N. Q. Azzahra, M. N. Pratama, E. A. Prionggo, T. Nurillatiffah, A. A. Pravitasari, dan F. Indrayatna, “Pengelompokan Tingkat Pengangguran, Kemiskinan, Dan Pendapatan Di Kabupaten/Kota Provinsi Jawa Barat,” 2023.

S. S. Rachmasari dan A. Kudus, “Perbandingan Penerapan Algoritme K-Means dan Fuzzy C-Means untuk Mengelompokkan Data Kinerja Dosen Universitas Islam Bandung,” 2021.

K. Ishak, “Faktor-Faktor Yang Mempengaruhi Pengangguran Dan Inflikasinyaterhadap Indek Pembangunan Di Indonesia,” Iqtishaduna: Jurnal Ilmiah Ekonomi Kita, vol. 7, no. 1, hlm. 22–38, 2018.

BPS, “Infografis-IPM-1-Desember-2023-ind.jpg (497×702).” Diakses: 7 Februari 2024. [Daring]. Tersedia pada: https://jatim.bps.go.id/backend/images/Infografis-IPM-1-Desember-2023-ind.jpg

BPS, “Tingat Partisipasi Angkatan Kerja (TPAK) - Tabel Statistik.” Diakses: 15 Januari 2025. [Daring]. Tersedia pada: https://gianyarkab.bps.go.id/id/statistics-table/2/NDIjMg==/tingat-partisipasi-angkatan-kerja-tpak-.html

A. Mardiana, “UMK Adalah Upah Minimum Kabupaten dan Kota, Ini Penjelasannya - Istilah Ekonomi Katadata.co.id.” Diakses: 25 Februari 2024. [Daring]. Tersedia pada: https://katadata.co.id/ekonopedia/istilah-ekonomi/656da1492983f/umk-adalah-upah-minimum-kabupaten-dan-kota-ini-penjelasannya

I. Permana dan F. N. S. Salisah, “Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation: The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm,” IJIRSE, vol. 2, no. 1, hlm. 67–72, Mar 2022, doi: 10.57152/ijirse.v2i1.311.

A. Asroni, H. Fitri, dan E. Prasetyo, “Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik),” st, vol. 21, no. 1, 2018, doi: 10.18196/st.211211.

F. Zubedi, “Analisis Cluster Hasil Try Out Siswa MTS AlHuda Gorontalo dengan Chi-Sim Cosimilarity dan K-Means Clustering,” jmpm j. mat. dan pendidik. mat., vol. 5, no. 1, hlm. 1–11, Mar 2020, doi: 10.26594/jmpm.v4i2.1706.

N. Qomariasih, “Implementasi K-Means Clustering Analysis untuk Mengelompokkan Kelurahan-Kelurahan Di DKI Jakarta Berdasarkan Jumlah Positif Covid-19.,” SLJIL, vol. 2, no. 07, hlm. 1003–1011, Jul 2021, doi: 10.46799/jst.v2i7.336.

S. Mashfuufah dan D. Istiawan, “Penerapan Partition Entropy Index, Partition Coefficient Index Dan Xie Beniindex Untuk Penentuan Jumlah Klaster Optimal Pada Algoritma

Fuzzy C-Means Dalam Pemetaan Tingkat Kesejahteraan Penduduk Jawa Tengah," Prosiding University Research Colloquium, hlm. 51–60, 2018.

D. Nurmin, M. N. Hayati, dan R. Goejantoro, "Penerapan Metode Fuzzy C-Means Pada Pengelompokkan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020," Jurnal EKSPONENSIAL, vol. 13, no. 2, hlm. 129–196, 2020.

B. Mailien, A. Salma, dan D. Fitria, "Comparison K-Means and Fuzzy C-Means Methods to Grouping Human Development Index Indicators in Indonesia," vol. 1, no. 1, hlm. 23–30, 2023.

A. Şenol, "VIASCKDE Index: A Novel Internal Cluster Validity Index for Arbitrary-Shaped Clusters Based on the Kernel Density Estimation," Computational Intelligence and Neuroscience, vol. 2022, hlm. 1–20, Jun 2022, doi: 10.1155/2022/4059302.

Oktoriandi, Dilla. 2022. "Penerapan Uji Q Cochran Terhadap Atribut Produk Laptop Menggunakan Multiple Response Analysis (MRA)." *Jurnal Riset Statistika* 1(2):127–34. doi:10.29313/jrs.v1i2.521.