

ANALISA PERBANDINGAN PENGELOMPOKKAN CURAH HUJAN 15 HARIAN PROVINSI DIY MENGGUNAKAN FUZZY CLUSTERING DAN K-MEANS CLUSTERING

Joko Suryanto¹

¹Program Studi Teknik Pertanian, STIPER Kutai Timur
Jl. Soekarno-Hatta No. 01 Sangatta Kutai Timur Kalimantan Timur 75387, Indonesia.
E-Mail: djakasoerya@yahoo.co.id

ABSTRAK

Analisa Perbandingan Pengelompokkan Curah Hujan 15 Harian Provinsi DIY Menggunakan Fuzzy Clustering Dan K-means Clustering. Tujuan penelitian ini adalah membandingkan pengelompokkan curah hujan 15 harian antara metode fuzzy c-means (FCM) dengan K-means. Penelitian menggunakan data curah hujan harian dari 22 stasiun hujan di wilayah DIY periode 1985 hingga 2009. *Principal Component Analysis* (PCA) digunakan untuk mereduksi jumlah variabel curah hujan 15 harian. Penentuan jumlah cluster optimum menggunakan indeks Xie-Beni (XB). Rasio standar deviasi (S_w/S_b), sum squared error (SSE), dan Silhouette (Si) dari kedua metode tersebut dibandingkan untuk menentukan metode yang sesuai untuk pengelompokkan data hujan 15 harian. Hasil analisis PCA diperoleh 4 komponen utama (PC) pertama mampu menjelaskan 82.9% variansi data, variabel tersebut kemudian digunakan sebagai input clustering FCM dan K-means. Berdasarkan indeks XB diperoleh jumlah cluster optimum metode K-means 6 cluster sedangkan metode FCM 5 cluster. Rerata rasio S_w/S_b metode K-means lebih kecil yaitu 0,243 dibandingkan metode FCM sebesar 0,289. Nilai Silhouette dan SSE metode K-means 0.46 dan 76, sedangkan metode FCM 0.24 dan 254. Nilai indeks validasi tersebut mengindikasikan bahwa K-means lebih sesuai untuk pengelompokkan curah hujan 15 harian Provinsi DIY dibandingkan metode fuzzy c-means.

Kata kunci : Clustering, Curah hujan 15 harian, Fuzzy c-means, Index Xie-Beni, K-means.

ABSTRACT

A Comparative Analysis of Fuzzy Clustering and K-means Clustering for 15-day Rainfall Data in DIY Province. This study aimed to compare performance two clustering methods (fuzzy c-means and K-means) on the basis validity index. The study used daily rainfall data for twenty five years (1985 to 2009) from 22 stations covering the DIY Province. PCA is used to reduce number 15-day rainfall variables and transform into new variable. Four validity index clustering: Xie-Beni index (XB), sum squared error (SSE), silhouette (Si) and standard deviation ratio (S_w/S_b) are used for compare performance of two clustering methods. The optimal number of cluster is determined using XB index, and result SSE, Si and ratio (S_w/S_b) are compared to find appropriate clustering algorithms to 15-day rainfall data. The study results showed that the first 4 PCs explains more than 82 % of total variance, than are used for data input in fuzzy c-means and K-means algorithms. The optimal number clusters according XB index are six and five for K-means and FCM method, respectively. The average ratio S_w/S_b K-means method (0.243) was smaller than fuzzy c-means method (0.289). The Silhouette and SSE index are 0.46 and 76, 0.24 and 254 for K-means and FCM respectively. The result indicated that K-means method to be better than fuzzy c-means for clustering 15-day rainfall data in DIY Province.

Key words : 15-day rainfall, Clustering, Fuzzy c-means, K-means, Xie-Beni index.

1. PENDAHULUAN

Curah hujan sebagai salah satu komponen iklim, sangat penting dalam bidang pertanian sebagai sumber air yang

dibutuhkan tanaman untuk memproduksi biomasa. Curah hujan satu wilayah dengan wilayah lain sangat bervariasi karena dipengaruhi oleh faktor lokal maupun global. Dalam skala lokal, posisi

geografis dan topografi sangat mempengaruhi variasi curah hujan secara spasial maupun temporal. Peningkatan produksi tanaman pangan melalui pemanfaatan data iklim, tidak terlepas dari keberhasilan dalam menginterpretasikan data iklim, terutama curah hujan yang sangat bervariasi. Interpretasi data curah hujan tersebut berkaitan dengan pengambilan keputusan manajemen pertanian, seperti pola tanam dan tata tanam, pengaturan irigasi (Tjasyono, 2004).

Salah satu usaha menginterpretasikan data hujan adalah menyederhanakan data dengan cara menetapkan wilayah-wilayah iklim yang homogen. Pewilayahan hujan dapat diartikan mendeliniasi stasiun curah hujan pada suatu wilayah ke dalam kelompok yang homogen. Homogen dalam bidang hidrologi diartikan sebagai wilayah yang mempunyai kesamaan karakteristik hidrologis, seperti geografis, fisik, statistik atau ciri stokastik (Modarres, 2006). Metode yang umum digunakan untuk pewilayahan fenomena hidrologi adalah analisis cluster. Analisis cluster merupakan sebuah teknik yang digunakan untuk mengklasifikasi beberapa objek menjadi wilayah-wilayah yang secara relatif bersifat homogen yang disebut cluster (Goncalves, *dkk.*, 2016).

Terdapat tiga jenis algoritma clustering; 1) hierarki, 2) non-hierarki, dan 3) gabungan antar 2 metode sebelumnya (hybrid). Termasuk metode hierarki adalah single linkage, complete linkage, average linkage, dan Ward, termasuk dalam metode non-hierarki yaitu fuzzy c-means dan K-means. Metode K-means adalah teknik algoritma untuk mengelompokkan item atau subyek menjadi K cluster dengan cara meminimalkan *sum of square* jarak dengan centroid cluster (Gudono, 2016). Algoritma K-means sangat sederhana dan dapat digunakan untuk

berbagai tipe data, namun beberapa kelemahan metode ini antara lain; metode ini selalu konvergen, sangat sensitif terhadap penentuan centroid awal sehingga hasil cluster sangat rentan terhadap solusi yang bersifat *local minimum* (Morisette dan Chartier, 2013).

Pengelompokkan dengan menggunakan algoritma fuzzy c-means (FCM) merupakan hasil perkembangan metode pengelompokkan non-hierarki dengan menggunakan logika fuzzy. FCM pertama kali diperkenalkan oleh Dunn tahun 1973 dan dikembangkan oleh Bezdek pada tahun 1981 yang digunakan untuk pengenalan pola (pattern recognition). Metode FCM sering digunakan dalam pengelompokkan karena metode ini memberikan hasil yang halus karena pembobotan dilakukan menggunakan himpunan fuzzy dan cukup efektif untuk meningkatkan homogenitas setiap cluster yang dihasilkan (Shihab, 2000 dalam Widodo dan Purnadi, 2013). Kelemahan metode FCM menurut Tian, *dkk.*, (2014) antara lain penentuan matriks partisi awal sangat sulit, dan keberadaan data outlier akan mempengaruhi hasil clustering.

Penggunaan metode fuzzy c-means dan K-means telah banyak dilakukan di bidang hidro-klimatologi. Puspitasari dan Haviluddin (2016) menggunakan metode K-means untuk pengelompokkan curah hujan tahunan pada 13 stasiun hujan yang tersebar di Kalimantan Timur dan diperoleh hasil *sum square error* (SSE) yang lebih rendah pada tiga cluster dibandingkan dua cluster. Sari (2014) menggunakan metode fuzzy c-means untuk mengelompokkan curah hujan bulanan di kota Bengkulu. Junaeni, *dkk.*, (2010) menyimpulkan bahwa metode *fuzzy clustering* lebih baik dibandingkan metode Ward untuk pengelompokkan curah hujan bulanan dari satelite *Tropiccal Rainfall Measuring Mission* (TRMM) 3B43 di

Kalimantan Barat. Pengelompokan terhadap 188 stasiun hujan di Turki berdasarkan curah hujan tahunan yang dilakukan Dikbas, *dkk.*, (2012) menggunakan FCM dengan data inputan koefisien variansi, posisi geografis dan tinggi tempat stasiun hujan diperoleh 6 kelompok stasiun hujan. Haqiqi dan Kurniawan (2015) menyimpulkan bahwa metode SFCM memberikan hasil indeks validasi yang lebih baik dibandingkan metode FCM pada jumlah cluster yang sama.

Penelitian ini bertujuan untuk membandingkan penggunaan metode clustering antara FCM dan K-means dalam mengelompokkan curah hujan 15 harian di Provinsi Daerah Istimewa Yogyakarta (DIY) ke dalam kelompok wilayah hujan yang mempunyai karakteristik yang sama. Secara fisiografis, Provinsi DIY terbagi menjadi satuan fisiografis yaitu Gunung Merapi, Pegunungan Kapur Selatan, Pegunungan Kulon Progo, dan dataran rendah yang berbatasan langsung dengan Samudera Hindia. Berdasarkan kondisi fisiografis tersebut, maka curah hujan yang diterima juga akan bervariasi. Sehingga dalam usaha pemanfaatan data curah hujan diperlukan analisa clustering untuk menyajikan informasi untuk pengambilan keputusan dalam kegiatan manajemen pertanian.

2. METODA PENELITIAN

2.1. Tempat dan Waktu

Penelitian dilaksanakan Provinsi Daerah Istimewa Yogyakarta (DIY) yang terletak di bagian selatan pulau Jawa bagian tengah, dan berbatasan dengan Provinsi Jawa Tengah dan Samudera Hindia. Secara geografis Provinsi DIY terletak pada 8° 30' – 7° 20' LS dan 109° 40' – 111° 0' BT. Provinsi DIY terbagi ke dalam 4

Kabupaten dan 1 Kotamadya, yaitu Kabupaten Sleman, Kabupaten Bantul, Kabupaten Kulonprogo, Kabupaten Gunung Kidul dan Kota Yogyakarta. Pada bulan Februari 2017.

2.2. Bahan dan Alat

Bahan yang digunakan berupa peta RBI Daerah Istimewa Yogyakarta yang diperoleh dari www.Bakosurtanal.go.id dan data curah hujan harian di wilayah Provinsi Daerah Istimewa Yogyakarta (DIY) periode tahun 1985 – 2009 yang diperoleh dari Balai Besar Wilayah Sungai (BBWS) Serayu Opak, Yogyakarta. Alat yang digunakan berupa *software* komputer program Microsoft Office, ArcMap 10.1 dan MATLAB R2014.

2.3. Analisa Data

Analisa Komponen Utama

Proses pembentukan komponen utama dilakukan dengan bantuan *software* Matlab R2014. Tahapan pembentukan komponen utama yang dilakukan adalah sebagai berikut:

1. Matriks X merupakan data pengamatan curah hujan 15 harian yang berukuran $n \times p$, dengan n adalah stasiun curah hujan, p adalah curah hujan 15 harian (mm).

$$X_{np} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

2. Melakukan normalisasi data awal untuk menghindari dominasi satu atau dua variabel menggunakan persamaan:

$$x = \frac{x_i - \bar{x}}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}}$$

3. Menentukan matriks kovarian dari matriks x .
4. Menentukan eigenvalue dari matriks kovarian, misalkan: $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$

5. Menentukan vektor eigen matriks kovarian, misalkan: $v_j = v_1, v_2, v_3, \dots, v_p$
6. Menghitung komponen utama (PC_j) yang dibentuk dari vektor eigen dan data yang telah ternormalisasi (x).

$$PC_p = x_j v_{jp} = x_1 v_{1p} + x_2 v_{2p} + x_3 v_{3p} + \dots + x_p v_{pp}$$
7. Menghitung variansi dan variansi kumulatif dari komponen utama.
8. Jumlah komponen utama yang akan digunakan dalam analisis cluster sebagai pengganti variabel awal ditentukan berdasarkan prosentase variansi kumulatif minimal 80 % (Rencher, 2001., dalam Junaeni, dkk., 2010).

Tahapan pengelompokkan Fuzzy C-means (FCM) sebagai berikut (Kusumadewi dan Hartati, 2006):

1. Menentukan input data berupa matriks berukuran n x m, dengan n = jumlah data yang akan dicluster, m = atribut setiap data. Menentukan jumlah cluster ($2 \leq c \leq n$), pangkat pembobot (w), iterasi maksimum, error terkecil yang diharapkan (ξ), dan iterasi awal (t=1).
2. Membangkitkan bilangan random μ_{ik} , $i=1,2,3,\dots,n$; $k=1,2,3,\dots,c$; sebagai elemen-elemen matriks partisi U.
3. Menghitung pusat cluster ke-k v_{kj} dengan $k= 1,2, \dots, c$; dan $j = 1,2,\dots,m$

$$v_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w x_{ij}}{\sum_{i=1}^n (\mu_{ik})^w}$$

4. Memperbarui keanggotaan fuzzy :

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{w-1}} \right]^{-1}$$

dengan:

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^m (x_{kj} - v_{ij})^2 \right]^{1/2}$$

5. Menentukan kriteria berhenti, yaitu perubahan matriks partisi pada iterasi

sekarang dan sebelumnya, yaitu $\Delta = \|U^t - U^{t-1}\|$. Apabila $\Delta \leq \xi$, maka iterasi dihentikan, namun apabila $\Delta > \xi$, maka menaikkan iterasi ($t = t + 1$), dan mengulangi langkah ke-3.

Dalam penelitian ini, pengelompokkan dengan metode Fuzzy C-means dilakukan dengan bantuan software Matlab R2014.

Tahapan pengelompokkan K-means adalah sebagai berikut:

Algoritma K-means adalah menghitung jarak setiap point dengan pusat cluster secara iteratif dengan tujuan meminimumkan fungsi objektif berupa *squared error* (Cebeci dan Yildiz, 2015). Algoritma K-means berbasis algoritma iteratif adalah sebagai berikut:

1. Menentukan pusat cluster c dari data x secara random.
2. Menghitung jarak setiap poin data terhadap pusat cluster menggunakan metode *euclidian distance* pada persamaan:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - k_{cj})^2}$$

3. Menentukan setiap poin data pada cluster dengan jarak ke pusat cluster terdekat.
4. Memperbarui pusat cluster menggunakan persamaan:

$$v = \sum_{i=1}^n x_i / n; i = 1,2,3 \dots n$$

5. Menghitung jarak setiap data terhadap pusat cluster baru dengan persamaan 1.
6. Mengulangi langkah 3 hingga 5 apabila data poin berada pada posisi cluster yang berbeda, dan jika posisi data poin pada cluster tidak berubah maka algoritma dihentikan.

Validasi Cluster

Indeks validasi cluster selain berfungsi untuk menentukan jumlah cluster optimal, juga dapat digunakan untuk menentukan algoritma clustering yang sesuai dengan data (Xie, dkk.,2011). Dalam penelitian ini, jumlah kelompok optimal ditentukan menggunakan indeks validitas Xie-Beni (XB), sedangkan metode yang sesuai untuk pengelompokan data hujan 15 harian ditentukan berdasarkan indeks *sum squared error* (SSE), koefisien Silhouette (Si) dan rasio standar deviasi (S_w/S_b).

1. Xie and Beni's Index (XB)

$$XB(c) = \frac{\sum_{i=1}^c \sum_{k=1}^N (u_{ik})^m \|x_k - v_i\|_{w_j}^2}{N \min_{ik} \|v_k - \bar{v}\|^2} = \frac{1}{c} \sum_{k=1}^c S_{(k)j}$$

dimana:

$$S_{(k)j} = \sqrt{\frac{\sum_{i=1}^n (x_{i(k)j} - \bar{x}_{(k)j})^2}{n - 1}}$$

$$S_{bj} = \left[\frac{1}{c - 1} \sum_{k=1}^c (\bar{x}_{(k)j} - \bar{x})^2 \right]^{1/2}$$

$$R = \frac{S_{w_{total}}}{S_{b_{total}}}$$

Jumlah cluster optimal terdapat pada nilai indeks XB minimum pada lembah pertama (Pravitasari, 2009; Widodo dan Purhadi, 2013; Rahmatika, dkk., 2015; Rizal dan Hakim, 2015).

2. Silhouette Coefficient (Si)

Indeks Silhouette (Si) mempunyai nilai $-1 \leq Si \leq 1$. Nilai Si positif berarti suatu titik data telah berada pada cluster yang tepat, sedangkan bernilai negatif menandakan adanya overlapping sehingga titik berada di cluster yang tidak tepat. Jika nilai Si = 0, maka titik berada di antara dua cluster (Susanto, 2016).

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}$$

$$s(i) = \begin{cases} 1 - a(i)/b(i) & \text{jika } a(i) < b(i) \\ 0 & \text{jika } a(i) = b(i) \\ a(i)/b(i) - 1 & \text{jika } a(i) > b(i) \end{cases}$$

3. Sum Square Error (SSE)

$$SSE(x, \Pi) = \sum_{i=1}^K \sum_{x_j \in c_i} \|x_j - m_i\|^2$$

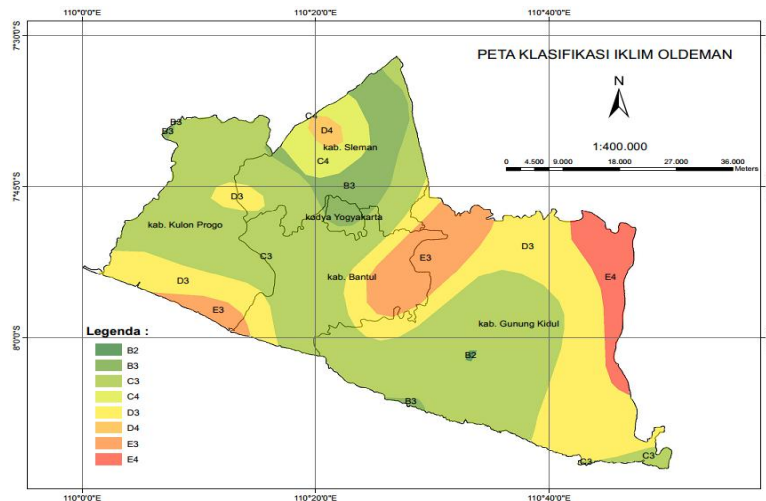
4. Rasio standar deviasi (R)

Rasio standar deviasi (R) telah berhasil digunakan oleh Ningsih, dkk., (2016) untuk menentukan metode terbaik diantara metode complete linkage dan average linkage. Nilai R merupakan perbandingan antara standar deviasi dalam kelompok (S_w) dengan standar deviasi antar kelompok (S_b), nilai R minimal menunjukkan algoritma metode cluster terbaik. Persamaan berikut digunakan untuk menghitung nilai S_w dan S_b :

3. HASIL PENELITIAN DAN PEMBAHASAN

3.1. Iklim Lokasi Penelitian

Tipe iklim di DIY berdasarkan metode Oldeman menggunakan data curah hujan bulanan selama 25 tahun (1985 – 2009) didominasi oleh tipe iklim C3 (Gambar 1.) dengan bulan basah dan bulan kering berurutan 5 – 6. Wilayah dengan tipe iklim C3 hanya dapat membudidayakan pada sawah satu kali musim tanam, sehingga budidaya padi untuk dua musim tanam berikutnya sangat bergantung pada adanya air irigasi.



Gambar 1. Peta klasifikasi iklim DIY

Rerata curah hujan 15 harian 22 stasiun hujan di DIY diperoleh hujan maksimum 196 mm terjadi pada periode 15 harian Januari 2, sedangkan terendah 5 mm terjadi pada periode Agustus 1 (Tabel 1.). Variasi curah hujan terendah

terjadi pada periode September 1 dengan nilai standar deviasi 3 mm, sedangkan variasi hujan tertinggi pada periode Februari 1 dengan nilai standar deviasi 41 mm.

Tabel. 1. Deskripsi curah hujan 15 harian stasiun hujan Provinsi DIY.

No	periode	mak	min	\bar{r}	sd	No	periode	mak	min	\bar{r}	sd
1	Jan1	219	125	161	28	13	Jul1	23	6	13	5
2	Jan2	261	137	196	32	14	Jul2	21	4	9	5
3	Peb1	275	118	189	41	15	Agst1	18	1	5	4
4	Peb2	214	104	151	31	16	Agst2	23	4	12	5
5	Mar1	202	83	146	30	17	Sep1	11	1	6	3
6	Mar2	191	93	140	29	18	Sep2	35	2	9	7
7	Apl1	185	59	112	40	19	Okt1	63	12	26	11
8	Apl2	129	35	66	24	20	Okt2	116	32	67	23
9	Mei1	87	15	44	21	21	Nov1	133	54	84	23
10	Mei2	71	19	35	13	22	Nov2	203	60	134	38
11	Jun1	56	18	34	11	23	Des1	190	84	142	30
12	Jun2	39	12	24	7	24	Des2	205	94	147	27

Tabel 2. menunjukkan karakteristik statistik data curah hujan 15 harian di DIY sepanjang 25 tahun (1985 – 2009) di 22 stasiun hujan. Curah hujan tahunan tertinggi 2592 mm/th di stasiun hujan Kempt Kabupaten Sleman, sedangkan rata-rata curah hujan tahunan terendah 1421 mm/th di stasiun hujan Ngawen Kabupaten Gunung Kidul. Rerata curah

hujan 15 harian maksimum adalah 330 mm, rata-rata tertinggi terjadi di stasiun hujan Panggang yaitu 450 mm dan terendah 238 mm di stasiun hujan Ngawen. Variasi hujan tertinggi terjadi di stasiun hujan Santan Kabupaten Sleman dengan nilai standar deviasi 121 mm sedangkan terendah 67 mm di stasiun hujan Ngawen.

Tabel. 2. Statistik deskriptif hujan 15 harian distasiun hujan DIY.

No	Stasiun Hujan	Bujur	Lintang	mm/th	rerata mm	maks mm	min mm	Sd mm
1	angin-angin	110.371	7.674	1706	73	272	5	79
2	beran	110.358	7.731	2549	106	380	6	110
3	bronggang	110.460	7.664	2345	98	347	6	102
4	dogongan	110.378	7.953	1645	66	341	1	90
5	gedangan	110.682	7.944	1704	72	307	2	86
6	gembongan	110.211	7.857	1765	75	291	2	90
7	godean	110.293	7.767	2281	94	338	5	101
8	kalibawang	110.264	7.676	2177	112	404	11	111
9	kalijoho	110.235	7.823	1817	75	299	2	89
10	kedungkeris	110.600	7.888	1595	70	287	2	84
11	kemptut	110.405	7.639	2592	109	385	7	113
12	kenteng	110.255	7.786	1611	68	298	1	84
13	ngawen	110.699	7.838	1421	59	238	4	67
14	panggang	110.346	7.844	2241	92	454	7	119
15	prumpung	110.392	7.707	2220	93	350	6	100
16	pundong	110.341	7.958	1718	73	330	2	94
17	sanden	110.272	7.975	1832	70	293	2	86
18	santan	110.416	7.787	2438	102	407	3	121
19	sapon	110.254	7.923	1823	75	319	3	92
20	seyegan	110.309	7.719	2192	92	309	4	96
21	tanjungtirto	110.463	7.794	1762	75	311	4	87
22	terong	110.452	7.891	1621	56	295	1	79

3.2. Analisa Komponen Utama

Analisa komponen utama (PCA) digunakan untuk mereduksi matriks data awal yang berukuran 22 x 24 ke matriks data baru yang berukuran 22 x 4 berdasarkan prosentase kumulatif varian. Empat komponen utama pertama mempunyai prosentase kumulatif varian sebesar 0.829, yaitu PC1 sebesar 0.598, PC2 0.689, PC3 0.768 dan PC4 0.829, dan eigenvalue lebih > 1 (Tabel 3.). Komponen utama tersebut sudah

memenuhi batas minimal prosentase kumulatif varian sebesar 0.80 (Rencher, 2001 dalam Junaeni, dkk., 2010). Sedangkan komponen utama yang tidak digunakan untuk analisa pengelompokkan mempunyai eigenvalue < 1. Reduksi dimensi data curah hujan menggunakan metode PCA ini telah banyak dilakukan oleh peneliti sebelumnya diantaranya oleh Junaeni (2014).

Tabel. 3. Hasil analisis komponen utama data curah hujan 15 harian.

PC	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Eigenvalue	14.350	2.194	1.882	1.461	0.847	0.686	0.604	0.388	0.367	0.275
Proportion	0.598	0.091	0.079	0.061	0.035	0.029	0.025	0.016	0.015	0.011
Cumulative	0.598	0.689	0.768	0.829	0.864	0.893	0.918	0.934	0.949	0.961

3.3. Analisa Pengelompokkan

Pengelompokkan stasiun curah hujan di DIY menggunakan 2 metode yaitu K-means dan Fuzzy C-means (FCM) dan sebagai data inputan adalah hasil dari analisis komponen utama atau *principal component analysis* (PCA) yaitu empat *score principal component*. Jumlah cluster ditentukan $c=2$ hingga

$c=10$ cluster untuk metode K-means maupun FCM. Pengelompokkan metode FCM menggunakan parameter yang tetap untuk setiap jumlah cluster yaitu nilai *default* pada matlab diantaranya; pangkat pembobot (w) 2, maksimum iterasi 100, dan error terkecil yang diharapkan (ξ) 10^{-5} .

Tabel. 4. Indeks validasi clustering metode K-means dan Fuzzy C-means.

cluster	K-means				FCM			
	Xie-Beni	Si	S_w/S_b	SSE	Xie-Beni	Si	S_w/S_b	SSE
2	0.168	0.69	0.24	171	0.152	0.69	0.24	1331
3	0.383	0.65	0.33	128	0.195	0.65	0.23	533
4	1.290	0.47	0.60	106	0.686	0.35	0.58	112
5	1.639	0.45	0.46	89	0.502	0.24	0.66	254
6	0.792	0.46	0.48	76	0.516	0.24	0.67	100
7	1.614	0.46	0.39	59	0.415	0.19	0.91	114
8	2.089	0.49	0.39	48	0.322	0.37	0.36	67
9	1.871	0.50	0.38	42	0.235	0.41	0.38	71
10	2.120	0.57	0.28	29	0.370	0.42	0.37	61

Tabel 4. memperlihatkan nilai indeks Xie-Beni (XB), *Silhouette Coefficient* (Si), rasio S_w/S_b dan *sum square error* (SSE). Indeks *Silhouette* menunjukkan ketepatan suatu point data (i) berada pada cluster k. Indeks *Si* mempunyai range antara -1 hingga 1, dimana nilai 1 menunjukkan point data i telah berada pada cluster yang tepat. Metode K-means maupun FCM memberikan kriteria validasi *Si* mendekati 1 pada jumlah cluster $c=2$ yaitu sebesar 0.69. Nilai minimum koefisien *Si* pada metode FCM terdapat pada $c=7$ yaitu 0.19, sedangkan minimum *Si* 0.45 pada metode K-means pada jumlah cluster $c=5$.

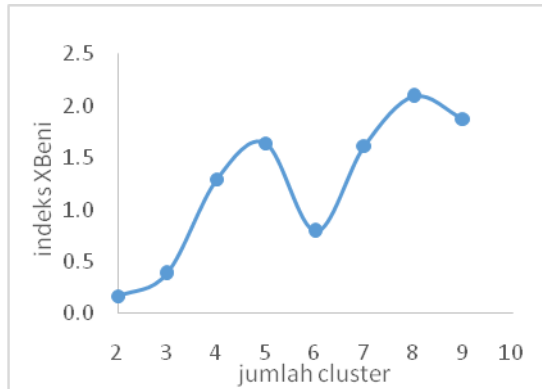
Nilai SSE dapat digunakan untuk validasi metode clustering karena meminimalkan SSE (kohesian) ekuivalen dengan memaksimalkan SSB (separation) (Nurwidodo dan Hariadi, 2013). Nilai SSE metode K-means maupun FCM cenderung menurun seiring dengan bertambahnya jumlah cluster, hingga SSE

minimum pada jumlah cluster $c=10$ sebesar 61 pada metode FCM dan 29 pada metode K-means. Berdasarkan kriteria SSE, baik pada metode K-means maupun FCM diperoleh jumlah cluster optimal pada $c=10$. Penurunan nilai SSE dapat dijelaskan bahwa pada saat pengelompokkan hanya terdiri dari satu cluster, maka SSE akan maksimum, dan pada saat satu cluster hanya terdiri dari satu anggota maka nilai SSE minimum karena jarak antara point data dengan pusat cluster = 0.

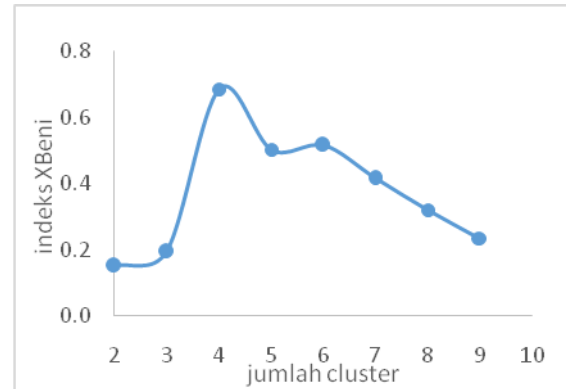
Berdasarkan Tabel 4. terlihat bahwa antara indeks validasi *Silhouette* dan SSE menunjukkan inkonsistensi dalam penentuan jumlah cluster optimal, yaitu . indeks *Si* mendekati 1 pada SSE maksimum. Menurut Thinsungnoen, *dkk.*, (2015) inkonsistensi validasi cluster antara indeks *Si* dan SSE disebabkan oleh data yang mengalami overlap, dan pada kondisi tersebut penggunaan kriteria SSE lebih mendekati kebenaran dibanding indeks *Silhouette*.

Inkonsistensi validasi cluster juga terjadi pada penggunaan kriteria rasio S_w/S_b . Pada metode K-means rasio S_w/S_b minimum pada nilai SSE maksimum yaitu 0.24, sedangkan pada metode FCM terdapat pada SSE jumlah cluster $c=3$. Adanya inkonsistensi indeks validasi

tersebut, maka dalam penelitian ini penentuan jumlah cluster optimal diputuskan menggunakan indeks Xie-Beni. Indeks ini mempunyai ketepatan dan keandalan yang tinggi dalam menentukan banyaknya kelompok optimum (Widodo dan Purhadi, 2013).



Gambar 2a. Indeks XB metode K-means



Gambar. 2b. Indeks XB metode FCM

Nilai indeks Xie-Beni pada Tabel 4. terlihat minimum pada jumlah cluster $c=2$, kemudian meningkat hingga cluster $c=10$. Jumlah cluster optimum ditentukan menggunakan kriteria indeks Xie-Beni minimum pada lembah pertama (Pravitasari, 2009; Rahmatika, dkk., 2015; Rizal dan Hakim, 2015). Metode K-means diperoleh indeks Xie-Beni 0.792 pada jumlah $c=6$, sedangkan pada

metode FCM diperoleh indeks Xie-Beni 0.502 jumlah $c=5$ (Gambar 2a. dan 2b.). Maka diperoleh jumlah cluster optimum metode K-means $c=6$ dan $c=5$ untuk metode FCM. Hasil pengelompokkan curah hujan 15 harian pada 22 stasiun hujan di DIY menggunakan metode K-means dan FCM diperlihatkan pada Tabel 5. berikut.

Tabel. 5. Hasil pengelompokkan metode K-means dan Fuzzy C-means.

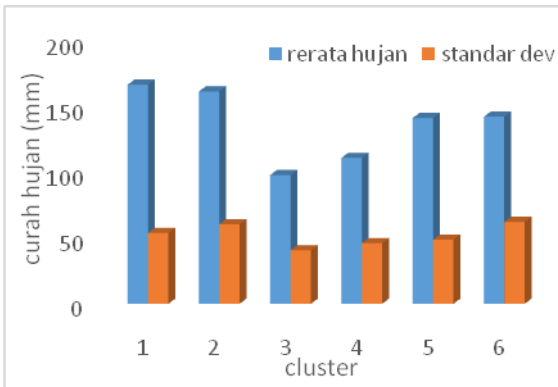
cluster	Stasiun hujan	
	K-means	Fuzzy C-means
1	kalibawang, kemptu	kalibawang
2	beran, santan	beran, bronggang, kemptu, panggang
3	angin-angin, gedangan, ngawen, terong	angin-angin, dogongan, gedangan, kedungkeris, kenteng, ngawen, tanjungtirto, terong
4	dogongan, gembongan, kalijoho, kedungkeris, kenteng, pundong, sanden, sapon, tanjungtirto	gembongan, kalijoho, pundong, sanden, sapon, seyegan
5	bronggang, godean, prumpung, seyegan	godean, prumpung, santan
6	panggang	

Gambar 3a. dan 3b. menunjukkan hasil cluster metode K-means dan fuzzy

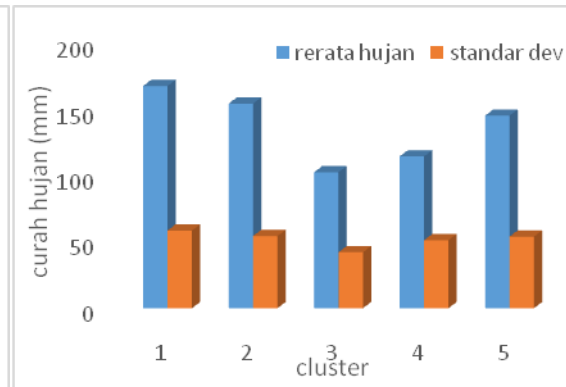
c-means terhadap data hujan 15 harian. Enam cluster yang diperoleh dari metode

K-means dapat terlihat bahwa cluster ke-1 adalah stasiun curah hujan yang mempunyai nilai rerata curah hujan pada musim hujan (Okt1 - Apl2) sebesar 168 mm. Rerata curah hujan pada musim hujan antara cluster ke-5 dan ke-6 menunjukkan kesamaan, namun yang

membedakan antara dua cluster tersebut adalah nilai standar deviasi yaitu 49 mm dan 63 mm pada cluster ke-6. Rerata curah hujan dan standar deviasi pada musim hujan paling rendah terdapat pada cluster ke-3 yaitu 98 mm dan 41 mm.



Gambar 3a. Hasil clustering metode K-means



Gambar. 3b. Hasil clustering metode FCM

Hasil pengelompokkan dikatakan baik apabila nilai standar deviasi intracuster (S_w) minimum sedangkan nilai antarcluster (S_b) maksimum. Tabel 6. memperlihatkan bahwa standar deviasi antarcluster (S_b) cenderung tetap, baik pada musim hujan (Nov1 – Apl1) maupun pada musim kemarau (Mei1 - Okt2), sedangkan nilai S_w cenderung lebih kecil pada musim kemarau. Karena

rasio S_w/S_b kecil pada musim kemarau, maka pengelompokkan curah hujan pada musim kemarau tersebut akan meningkatkan nilai validasi cluster dibandingkan curah hujan pada musim hujan yang disebabkan oleh variansi curah hujan yang lebih rendah pada musim kemarau dibandingkan pada musim hujan.

Tabel. 6. Nilai S_w dan S_b periode 15harian metode K-means dan Fuzzy C-means.

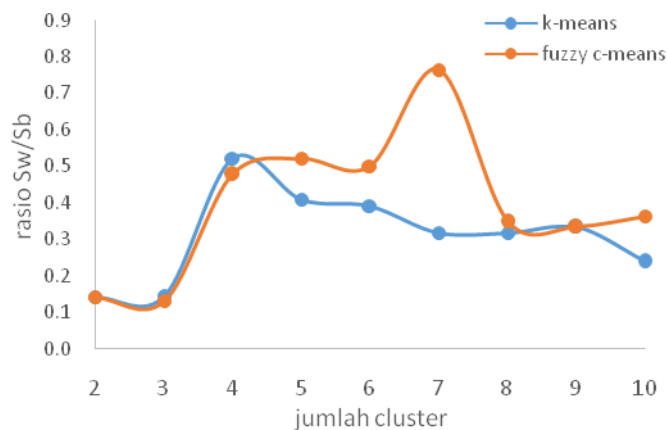
No	periode	K-means		FCM		No	periode	K-means		FCM	
		S_w	S_b	S_w	S_b			S_w	S_b	S_w	S_b
1	Jan1	10	32	12	29	13	Jul1	3	37	3	37
2	Jan2	18	39	20	35	14	Jul2	4	37	4	38
3	Peb1	21	46	20	46	15	Agst1	1	38	2	39
4	Peb2	14	34	16	35	16	Agst2	3	37	4	37
5	Mar1	13	33	16	33	17	Sep1	2	36	2	36
6	Mar2	10	33	14	33	18	Sep2	4	41	5	38
7	Apl1	14	44	19	48	19	Okt1	8	38	8	38
8	Apl2	11	40	12	45	20	Okt2	6	40	8	43
9	Mei1	6	43	9	45	21	Nov1	12	35	13	37
10	Mei2	8	37	8	40	22	Nov2	16	44	20	43
11	Jun1	6	36	7	37	23	Des1	14	30	14	32
12	Jun2	4	35	5	36	24	Des2	10	28	13	28

Tabel 7. menunjukkan nilai S_b pengelompokkan metode K-means maupun FCM tidak terdapat perbedaan yang signifikan, sedangkan nilai S_w sangat berbeda nyata. Hal tersebut ditunjukkan dengan hasil uji t-paired nilai S_b pada tingkat kepercayaan 99 %

menunjukkan nilai signifikansi $0.633 > 0.05$, sedangkan nilai S_w diperoleh signifikansi $0.02 < 0.05$. Metode K-means cenderung menghasilkan nilai S_w yang lebih kecil dibandingkan metode FCM pada jumlah cluster $c=5$, $c=6$, $c=7$, $c=8$, dan $c=10$.

Tabel. 7. Nilai S_w dan S_b metode K-means dan Fuzzy C-means.

	c2	c3*	c4*	c5	c6	c7	c8	c9	c10*	rerata
K-means										
Sw	14	12	12	9	9	7	7	7	5	9
Sb	99	83	23	22	23	22	22	21	21	37
Fuzzy C-means										
Sw	14	12	11	12	11	13	8	7	8	11
Sb	99	91	23	23	22	17	23	21	22	38



Gambar 4. Grafik rasio S_w/S_b terhadap jumlah cluster.

Gambar 4. dapat dilihat bahwa metode K-means menghasilkan rasio S_w/S_b yang lebih rendah dibandingkan FCM. Berdasarkan Tabel 4. rasio S_w/S_b metode K-means 0,243, sedangkan metode FCM sebesar 0,289. Berdasarkan perbandingan nilai S_w/S_b , maka diperoleh kesimpulan metode K-means lebih sesuai digunakan untuk pengelompokkan curah hujan 15 harian di DIY dibandingkan metode FCM. Nilai koefisien Silhouette metode FCM pada $c=5$ sebesar 0.24, lebih kecil dibandingkan metode K-means pada $c=6$ sebesar 0.46, Nilai SSE metode K-means 76, sedangkan pada

metode FCM 254. Kedua indeks validasi tersebut menguatkan kesimpulan bahwa metode K-means lebih sesuai untuk pengelompokkan hujan 15 harian dibandingkan metode FCCM. Sehingga diperoleh hasil pengelompokkan curah hujan 15 harian di DIY sebanyak 6 cluster.

4. KESIMPULAN

Dua metode clustering digunakan untuk mengelompokkan curah hujan 15 harian Provinsi DIY yaitu metode K-means dan fuzzy c-means. Jumlah cluster

optimal ditentukan menggunakan indeks Xie-Beni dan diperoleh jumlah cluster metode K-means 6 cluster sedangkan metode fuzzy c-means 5 cluster. Berdasarkan rasio S_w/S_b , metode K-means lebih baik dibandingkan metode fuzzy c-means, rerata rasio S_w/S_b metode K-means lebih kecil yaitu 0,243, sedangkan metode fuzzy c-means sebesar 0,289. Indeks validasi Silhouette dan SSE mempertegas kesimpulan bahwa metode K-means lebih sesuai untuk pengelompokan curah hujan 15 harian di Provinsi DIY.

DAFTAR PUSTAKA

- [1] Cebeci, Z. dan Yildiz, F. 2015. *Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures*. Journal of Agricultural Informatics. Vol. 6 No. 3 pp 13-23.
- [2] Dikbas, F., Firat, M., Koc, A.C., dan Gungor, M. 2012. *Classification of Precipitation Series Using Fuzzy Cluster Method*. Int. Journal of Climatology. 32. pp 1596 – 1603.
- [3] Goncalves, M.F., Blanco, C.J.C., Santos, V.,C., Oliveira, L.,L.,S., dan Pessoa, F.C.L. 2016. *Identification of Rainfall Homogenous Regions Taking into Account El-Nino and La-Nina and Rainfall Decrease in the State of Para, Brazilian Amazon*. Acta Scientiarum Technology. Vol.38. No.2. pp. 209 – 2016.
- [4] Gudono. 2016. *Analisis Data Multivariat*. Cetakan Kedua Edisi 4. BPF. Yogyakarta.
- [5] Haqiqi, B.N. dan Kurniawan, R. 2015. *Analisis Perbandingan Metode C-Means dan Subtractive Fuzzy C-Means*. Media Statistika. Vol.8 No. 2 pp 59-67.
- [6] Junaeni, Ina. 2014. *Dampak Penerapan Principal Component Anaysis (PCA) Dalam Clustering Curah Hujan di Pulau Jawa, Bali, Dan Lombok*. Jurnal Sains Dirgantara Vol. 11 no. 2. pp; 82-99.
- [7] Junaeni, I., Yuliani, D., Ayahbi, R., Noersomadi, Hardjana, T., dan Nurzaman. 2010. *Pengelompokan Wilayah Curah Hujan Kalimantan Barat Berbasis Metode Ward dan Fuzzy Clustering*. Jurnal Sains Dirgantara Vol. 7 no. 2. pp; 82-99
- [8] Kusumadewi, S., dan Hartati,. 2006. *Neuro Fuzzy: Integrasi Sistem Fuzzy dan Jaringan Syaraf*. Graha Ilmu. Yogyakarta.
- [9] Modarres, Reza.2006. *Regional Precipitation Climates of Iran*. Journal of Hydrology. Vol 45. No. 1. pp 15 – 29.
- [10]Morissette, L. dan Chartier, S.2013. *The K-means Clustering Technique: Consideration and Implementation in Mathematica*. Tutorials in Quantitative Methods for Psychology. Vol 9 No. 1. pp 15 – 24.

- [11] Ningsih, S., Sri, W., dan Nasution, Y.N. 2016. *Perbandingan Kinerja Metode Complete Linkage dan Average Linkage dalam Menentukan Hasil Analisis Cluster: Studi Kasus Produksi Palawija Provinsi Kalimantan Timur 2014/2015*. Prosiding Seminar Sains dan Teknologi FMIPA UNMUL. Vol. 1 No. 1 pp. 46 – 50.
- [12] Nurwidodo, W.R.S. dan Hariadi, M. 2013. *Sistem Pendukung Keputusan Multidimensi Menggunakan K-means Clustering Berbasis Mahalobis Distance*. Prosiding Seminar Nasional Manajemen Teknologi XVII. pp. 20-4–20-11.
- [13] Pravitasari, A.A. 2009. *Penentuan Banyaknya Kelompok dalam Fuzzy C-means Cluster Berdasarkan Proporsi Eigen Value Dari Matriks Similarity dan Indeks XB (Xie dan Beni)*. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika. FMIPA UNY. pp. 623 - 632.
- [14] Puspitasari, N. dan Havaluddin. 2016. *Penerapan Metode K-Means dalam Pengelompokan Curah Hujan di Kalimantan Timur*. Proceeding Seminar Nasional Riset Ilmu Komputer (SNRIK 2016) Universitas Muslim Indonesia. Makasar.
- [15] Rahmatika, L., Suparti, dan Safitri, D. 2015. *Analisis Kelompok Dengan Algoritma Fuzzy C-means dan Gustafson Kessel Clustering pada Indeks LQ45*. Jurnal Gaussian. Vol 4 No. 3. pp. 543 – 552.
- [16] Rizal, A. S. dan Hakim R.B. F. .2015. *Metode K-means Cluster dan Fuzzy C-means Cluster; Studi Kasus Indeks Pembangunan Manusia di Kawasan Indonesia Timur Tahun 2012*. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015. pp. 643 – 657.
- [17] Sari, H. L.2014. *Fuzzy Clustering Dalam Pengclusteran Data Curah Hujan Kota Bengkulu Dengan Algoritma C-Means*. Jurnal Ilmiah MATRIK. Vol 16. No. 2. pp. 115 – 124.
- [18] Susanto, E.B.2016. *Evaluasi Hasil Klaster pada Dataset Iris, Soybean-small, Wine menggunakan Algoritma Fuzzy C-means dan K-means++*. Surya Informatika. Vol.2 No.. pp 6 – 13.
- [19] Thinsungnoen, T., Kaoungku, N., dan Durongdumronchai, P. 2015. *The Clustering Validity with Silhouette and Sum Squared Error*. Proceedings of the 3rd International Conference on Industrial Application Engineering. The Institute of Industrial Applications Engineering. Japan. pp 44 – 51.
- [20] Tian, W., Zheng, Y., Yang, R., Ji, S., dan Wang, J., 2014. *Research on lustering based Meteorological Data Mining Methods*. Advance Science and Tecchnology Letters. Vol 179. pp 06 – 112.

- [21] Tjasyono, Bayong. 2004. *Klimatologi*. Edisi ke-2. ITB. Bandung.
- [22] Widodo, A., dan Purhadi. 2013. *Perbandingan Metode Fuzzy C-means Clustering dan Fuzzy C-shell Clustering; studi Kasus Kabupaten/Kota di Pulau Jawa Berdasarkan Variabel Pembentuk Indeks Pembangunan Manusia*. IndoMS Journal on Statistics. Vol. 1 No 2. pp. 61 – 71.
- [23] Xie, N., Hu, L., Luktarhan, N., dan Zhao, K. 2011. *A Clasification of Cluster Validity Indexes Based on Membership Degree and Applications*. WISM 2011, part I LNCS. pp. 43 – 50.