

**APLIKASI K-MEANS DAN FUZY CLUSTERING DALAM
PENGELOMPOKAN KECAMATAN DI KABUPATEN BANYUMAS**

Jajang

Universitas Jenderal Soedirman
jajang@unsoed.ac.id

Nunung Nurhayati

Universitas Jenderal Soedirman
nunung.nurhayati@unsoed.ac.id

Yhenis Apriliana

Universitas Jenderal Soedirman
yhenis.apriliana@mhs.unsoed.ac.id

ABSTRACT. Clustering N objects into c clusters can be used to get information about data observation. Among the clustering methods are K-Means (KMC) and Fuzzy C-means (FCM) methods. In the K-means method, objects are members or not members of the cluster, while in the FCM method, objects are included in the cluster based on the degree of membership. This study discusses the implementation of KMC and FCM in the clustering of sub-districts in Banyumas Regency based on total of population, the number of health workers and the number of health facilities and infrastructure. The results showed that the KMC and FCM methods produced the same cluster membership. Furthermore, the analysis of clustering based on the number of population, the number of health workers and the number of health facilities and infrastructure (scenario 1) and based on the number of health workers and the number of health facilities and infrastructure which have been corrected by population (scenario 2). The percentage of the variance ratio between clusters to the total variance in scenario 1 is 69% while in scenario 2 it is 85%. Clustering based on scenario 2 is better than scenario 1.

Keywords. *number of health workers, number of health facilities, K-Means, FCM.*

ABSTRAK. *Clustering* N objek menjadi c cluster dapat digunakan untuk mendapatkan informasi tentang data. Di antara metode *clustering* adalah metode K-Means (KMC) dan Fuzzy C-means (FCM). Pada metode metode K-means, objek sebagai anggota atau bukan anggota *cluster*, sedangkan pada metode FCM memasukan objek ke *cluster* berdasarkan derajat keanggotaan. Penelitian ini membahas implementasi KMC dan FCM dalam pengelompokan kecamatan-kecamatan di Kabupaten Banyumas berdasarkan jumlah penduduk, jumlah tenaga kesehatan dan jumlah sarana dan prasarana kesehatan. Hasil penelitian menunjukkan bahwa metode KMC dan FCM menghasilkan keanggotaan *cluster* yang sama. Lebih lanjut analisis pengelompokan dilakukan melalui berdasarkan jumlah penduduk, jumlah tenaga kesehatan dan jumlah sarana dan prasarana kesehatan (skenario 1) dan berdasarkan , jumlah tenaga kesehatan dan jumlah sarana dan prasarana kesehatan yang sudah dikoreksi dengan jumlah penduduk (scenario 2). Persentase rasio variansi antar *cluster* terhadap variansi total pada scenario 1 sebesar 69% sedangkan pada scenario 2 sebesar 85%. *Clustering* berdasarkan scenario 2 lebih baik daripada scenario 1.

Kata Kunci. *jumlah tenaga kesehatan, jumlah sarana kesehatan, K-Means, FCM.*

1. PENDAHULUAN

Menurut Pusparini (2012), penduduk adalah orang-orang yang berada di dalam suatu wilayah yang terikat oleh aturan-aturan yang berlaku dan saling berinteraksi satu sama lain secara terus-menerus. Jumlah penduduk akan terkait erat dengan indeks pembangunan manusia (IPM). Apabila sumber daya manusia meningkat, baik dari sisi pendidikan, kesehatan, dan perekonomian maka IPM sebuah wilayah akan meningkat pula. Terkait dengan masalah kesehatan, tentunya hal ini terkait erat dengan jumlah tenaga kesehatan dan sarana prasarana kesehatan. Menurut Undang-Undang No. 36 Tahun 2014, tenaga kesehatan terdiri dari tenaga medis (dokter dan dokter gigi), tenaga keperawatan (perawat dan bidan), tenaga kefarmasian (apoteker, analisis farmasi, dan asisten apoteker), tenaga kesehatan (tenaga kesehatan masyarakat, tenaga gizi, tenaga keterampilan fisik, dan tenaga keteknisian medis).

Adanya keterikatan satu dengan yang lainnya dalam menunjang kualitas sebuah wilayah, maka perlu untuk mengenal perkembangan sebuah wilayah, khususnya dalam penelitian ini, adalah kecamatan-kecamatan di Kabupaten Banyumas. Analisis *cluster* merupakan analisis yang tepat untuk kasus ini. Tokushig dkk (2007) telah mengembangkan algoritma *clustering* dari *crisp* and *fuzzy k-means* untuk analisis data variabel ganda (*multivariate*). *Clustering* telah terbukti cukup efektif dalam membuka struktur data–data yang terkait dengan kesehatan (Khanmohammadi dkk, 2017)

Berdasarkan uraian di atas, maka penelitian ini bertujuan untuk mengklasifikasikan kecamatan-kecamatan yang ada di kabupaten Banyumas menggunakan metode *fuzzy clustering mean* (FCM) dan *k-mean clustering* berdasarkan variabel jumlah penduduk, sarana dan prasarana kesehatan, dan tenaga kesehatan.

2. METODE PENELITIAN

3.1 Data

Data yang digunakan dalam penelitian ini adalah data jumlah penduduk, jumlah tenaga kesehatan, dan jumlah sarana kesehatan di Kabupaten Banyumas

tahun 2020 yang diperoleh dari Badan Pusat Statistik kabupaten Banyumas (BPS, 2020).

3.2 Metode *Custer K-Means*

Analisis *Cluster* merupakan salah satu teknik peubah ganda dalam statistika untuk mengkaji kemiripan dan ketidakmiripan (jarak) antar objek. Tujuan analisis *cluster* adalah untuk meminimumkan variansi di dalam satu *cluster* dan memaksimumkan variansi antar *cluster*.

Salah satu metode *clustering* adalah metode K-means. Dalam pembentukan *cluster*, metode ini didasarkan pada algoritma K-means yang menggunakan ukuran ketidakmiripan (jarak). erdapat beberapa ukuran jarak yang dapat digunakan, salah satunya adalah jarak Euclid (*Euclidean Distance*). Jarak Eulidean antara dua objek, o_i dan o_{i^*} , d_{ii^*} , didefinisikan sebagai

$$d_{ii^*} = \sqrt{\sum_{j=1}^k (o_{ij} - o_{i^*j})^2} \quad (1)$$

dengan o_{ij} nilai pengamatan i untuk variabel j , o_{i^*j} nilai pengamatan i^* untuk variabel j , dan d_{ii^*} jarak antara dua objek. Untuk lebih jelasnya, berikut adalah algoritma K-means :

- (1) menentukan banyaknya *cluster*, misalkan k *cluster*;
- (2) menentukan pusat *cluster*, untuk tahap awal dipilih secara acah;
- (3) hitung jarak setiap objek ke masing-masing pusat *cluster*, dan alokasikan masing-masing objek ke dalam *cluster* paling dekat, dan hitung pusat *cluster* yang baru;
- (4) ulangi langkah (2) sampai (3) sampai tidak ada lagi perpindahan objek;
- (5) menghitung variansi setiap *cluster*.

3.3 Himpunan Fuzzy

Konsep keanggotaan elemen x dalam himpunan tegas (*crisp*) A , $\mu_A(x)$, memiliki dua kemungkinan, yaitu 1 menunjukkan keanggotaan, dan 0 menunjukkan bukan anggota himpunan. Namun pada kenyataannya, terdapat beberapa himpunan yang tidak tepat bila menggunakan himpunan tegas.

Himpunan *fuzzy* digunakan untuk mengantisipasi hal tersebut. Dalam himpunan *fuzzy*, suatu elemen dapat masuk dalam dua himpunan yang berbeda. Seberapa besar eksistensinya dalam himpunan tersebut dapat dilihat pada derajat keanggotaannya.

Fungsi keanggotaan adalah fungsi yang memetakan semesta pembicaraan ke himpunan nilai keanggotaan (Sivanandam, 2007). Dalam teori himpunan *fuzzy*, keanggotaan suatu elemen di dalam himpunan dinyatakan dengan derajat keanggotaan (*membership values*) yang nilainya terletak di dalam selang $[0,1]$, yaitu $\mu_A: X \rightarrow [0,1]$.

3.4 Fuzzy C-means

Jika *cluster* dianggap sebagai sebuah himpunan bagian dari data, maka himpunan bagian tersebut adalah *fuzzy* atau *crisp*. Pada konsep *hard clustering* seperti *K-means*, nilai objek dapat dimasukkan ke dalam *cluster* (nilai keanggotaan 1) atau tidak (nilai keanggotaan 0). Pada *soft clustering*, seperti *Fuzzy C-means* (FCM) membolehkan objek kepunyaan beberapa *cluster* secara simultan dengan perbedaan tingkat keanggotaan. Misalkan c adalah banyaknya *cluster* yang sudah diketahui, maka tujuan dari *clustering* adalah mempartisi data X ke dalam c *cluster* (Abonyi, 2007).

Partisi *fuzzy* data X dapat direpresentasikan oleh matriks partisi *fuzzy* U berukuran $c \times N$, $= \{\mu_{i,k}\}_{c \times N}$, dengan $\mu_{i,k}$ adalah derajat keanggotaan observasi k milik *cluster* i , $\mu_{i,k} \in [0,1]$, $\sum_{i=1}^c \mu_{i,k} = 1, 1 \leq k \leq N$, dan $0 < \sum_{i=1}^N \mu_{i,k} < N, 1 \leq k \leq c$. Dari matriks partisi *fuzzy*, selanjutnya dapat dibentuk ruang partisi *fuzzy* $M_{fc} = \{U \in R^{c \times N} | \mu_{i,k} \in [0,1], \forall i, k; \sum_{i=1}^c \mu_{i,k} = 1, \forall k; 0 < \sum_{i=1}^N \mu_{i,k} < N, \forall i\}$.

Tujuan FCM adalah meminimumkan fungsi tujuan $J(X|U, V, \lambda)$, dimana $J(X|U, V, \lambda) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k})^m d_{i,kA}^2 + \sum_{k=1}^N \lambda_k (\sum_{i=1}^c \mu_{i,k} - 1)$. Nilai keanggotaan dan pusat *cluster* yang dapat meminimumkan fungsi $J(X|U, V, \lambda)$ adalah

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^c (d_{i,kA}^2 / d_{j,kA}^2)^{2/(m-1)}}, \text{ dan } v_i^l = \frac{\sum_{k=1}^N (\mu_{i,k}^{l-1})^m x_k}{\sum_{k=1}^N (\mu_{i,k}^{l-1})^m}, \text{ , } 1 \leq i \leq c, \quad (2)$$

Untuk mendapatkan hasil cluster dengan menggunakan metode FCM, dapat dilakukan dengan menggunakan algoritma *Fuzzy C-means* (FCM). Misalkan X adalah matriks data, c adalah banyak cluster c, 1 ≤ i ≤ c, bobot pangkat m > 1, batas toleransi, ε > 0, dan *norm-inducing matrix* A (dalam kasus jarak Euclid, A adalah matriks identitas). Algoritma *Fuzzy C-means* (FCM) adalah sebagai berikut:

- Inisialisasi matriks partisi secara acak $U^{(0)} \in M_{fc}$,
- untuk $l=1,2,3,\dots$, lakukan tahapan berikut :

(1) $v_i^l = \frac{\sum_{k=1}^N (\mu_{i,k}^{l-1})^m x_k}{\sum_{k=1}^N (\mu_{i,k}^{l-1})^m}, 1 \leq i \leq c$

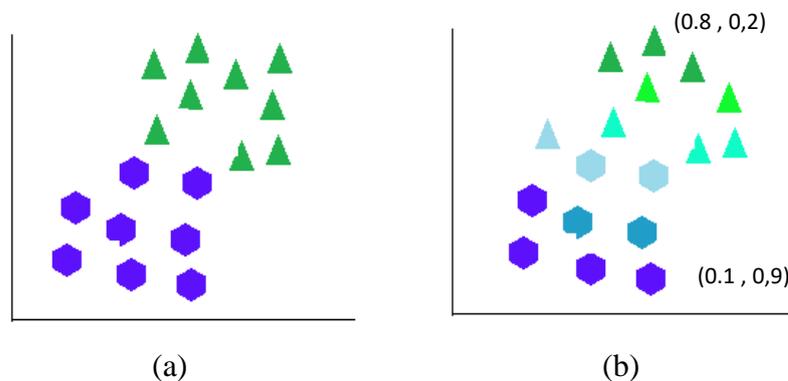
(2) Hitung jarak $d_{i,kA}^2 = (x_k - v_i)^T A (x_k - v_i), 1 \leq i \leq c, 1 \leq k \leq N$

(3) Perbaharui matriks U dengan memperbaharui $\mu_{i,k}$ dengan

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^c (d_{i,kA}^2 / d_{j,kA}^2)^{2/(m-1)}},$$

$$\|U^{(l)} - U^{(l-1)}\| < \varepsilon.$$

Gambar 1 mngilustrasikan perbedaan cluster KMC dan FCM. Misalkan gambar segitiga warna hijau mewakili cluster 1 dan bentuk segienam warna biru mewakili cluster 2.



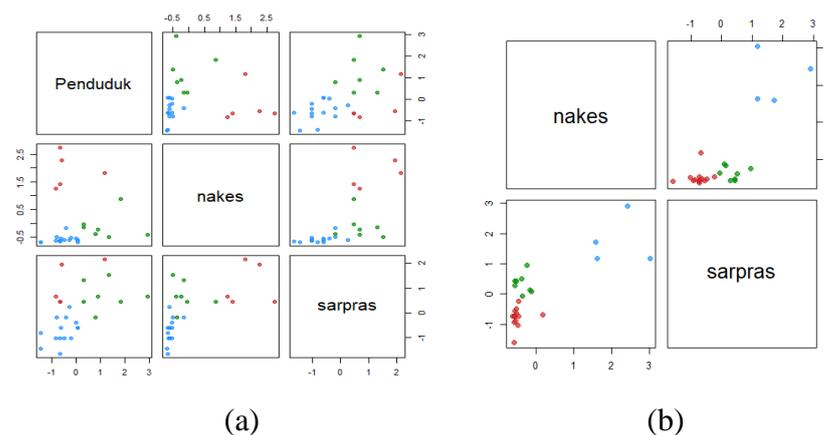
Gambar 1. Ilustrasi hasil Cluster KMC dan FCM

Pada Gambar 1.a, antar anggota cluster sangat jelas perbedaannya. Namun pada Gambar 1.b. terdapat nilai keanggotaan *cluster*. Misalkan untuk objek dengan nilai keanggotaan *cluster* 1 sebesar 0,8 dan nilai keanggotaan *cluster* 2 sebesar 0,2, (0,8,0,2), artinya derajat keanggotaan objek tersebut lebih dekat ke *cluster* 1. Sedangkan, untuk objek dengan nilai keanggotaan *cluster* 1 sebesar 0,1 dan nilai keanggotaan *cluster* 2 sebesar 0,9 (0,1, 0,9) artinya derajat keanggotaan objek tersebut lebih dekat ke *cluster* 2. *Cluster* yang baik adalah *cluster* yang menghasilkan variansi dalam *cluster* yang minimum, atau variansi antar *cluster* maksimum.

3. HASIL DAN PEMBAHASAN

3.1 Scatterplot

Analisis korelasi umum digunakan dalam tahapan analisis pemodelan. Tujuan dari analisis korelasi adalah untuk melihat gambaran visual mengenai keterkaitan linier antar variabel yang akan dianalisis. Dalam analisis *Cluster*, analisis korelasi dapat menjadi indikator bagaimana objek-objek atau individu mengelompok. Hasil analisis *scatterplot* kecamatan-kecamatan di kabupaten Banyumas disajikan pada Gambar 2.



Gambar 2. *Scatterplot* berdasarkan (a) semua variabel, dan (b) hanya berdasarkan jumlah tenaga kesehatan (*nakes*) dan sarana prasarana (*sarpras*)

3.2 Hasil *Cluster* Berdasarkan Semua Variabel

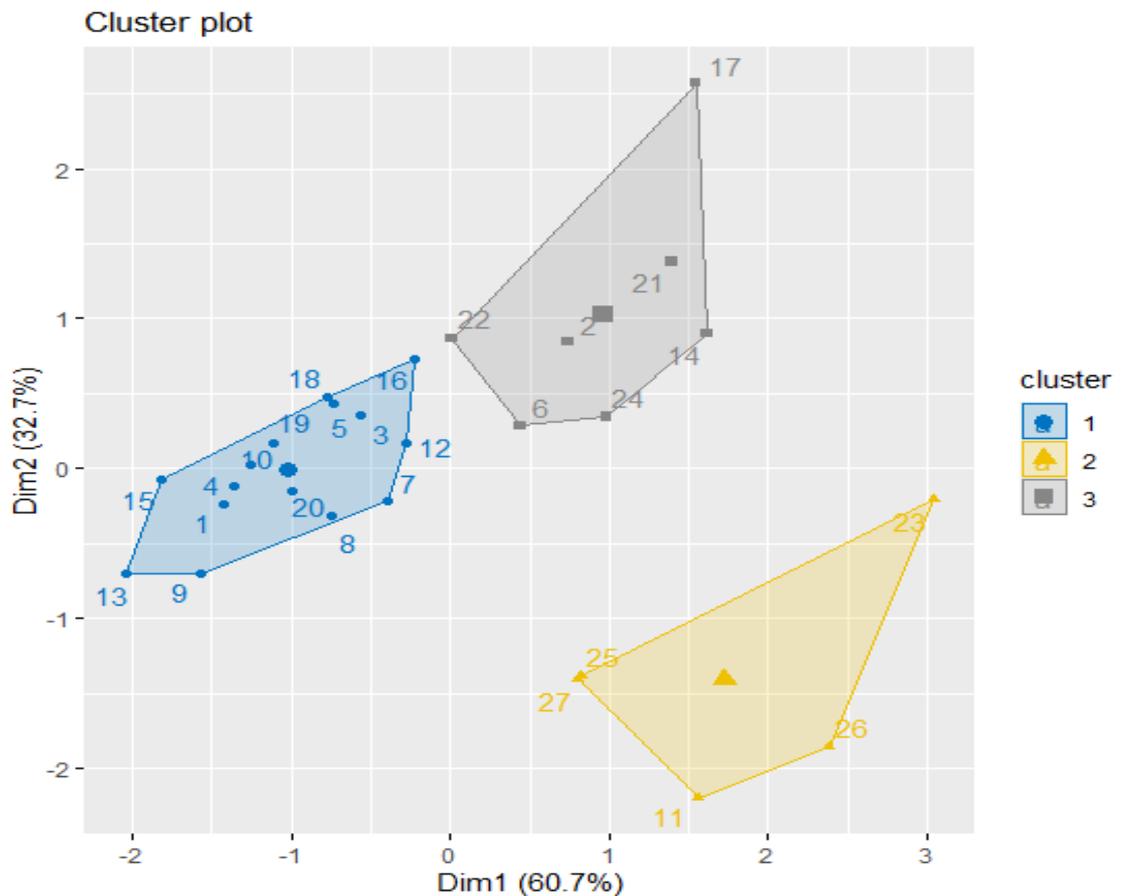
Hasil analisis *cluster* dengan menggunakan semua variabel disajikan pada Tabel 1. Berdasarkan Tabel 1, terlihat bahwa semua *cluster* dari hasil analisis *K-mean* dan FCM sama. Oleh karena itu, dalam hal ini kedua metode sama baiknya.

Tabel 1. *Cluster* berdasarkan semua variabel

no	Kecamatan	FCM			K-Means		
		1	2	3	1	2	3
1	Lumbir	0,958	0,017	0,025	1	0	0
2	Wangon	0,012	0,008	0,980	0	0	1
3	Jatilawang	0,827	0,044	0,130	1	0	0
4	Rawalo	0,973	0,011	0,017	1	0	0
5	Kebasen	0,872	0,034	0,094	1	0	0
6	Kemranjen	0,186	0,104	0,710	0	0	1
7	Sumpiuh	0,794	0,076	0,130	1	0	0
8	Tambak	0,862	0,055	0,083	1	0	0
9	Somagede	0,844	0,073	0,083	1	0	0
10	Kalibagor	0,979	0,008	0,014	1	0	0
11	Banyumas	0,077	0,837	0,086	0	1	0
12	Patikraja	0,612	0,101	0,287	1	0	0
13	Purwojati	0,811	0,089	0,101	1	0	0
14	Ajibarang	0,126	0,178	0,696	0	0	1
15	Gumelar	0,860	0,057	0,083	1	0	0
16	Pekuncen	0,479	0,079	0,442	1	0	0
17	Cilongok	0,177	0,148	0,675	0	0	1
18	Karanglewas	0,869	0,034	0,097	1	0	0
19	Kedungbanteng	0,954	0,016	0,031	1	0	0

20	Baturraden	0,987	0,005	0,008	1	0	0
21	Sumbang	0,084	0,080	0,836	0	0	1
22	Kembaran	0,304	0,080	0,616	0	0	1
23	Sokaraja	0,124	0,511	0,365	0	1	0
24	Purwokerto Selatan	0,127	0,141	0,732	0	0	1
25	PurwokertoBarat	0,071	0,858	0,072	0	1	0
26	Purwokerto Timur	0,066	0,831	0,103	0	1	0
27	Purwokerto Utara	0,082	0,835	0,083	0	1	0
Persentase variansi antar <i>cluster</i> dan total		69%					

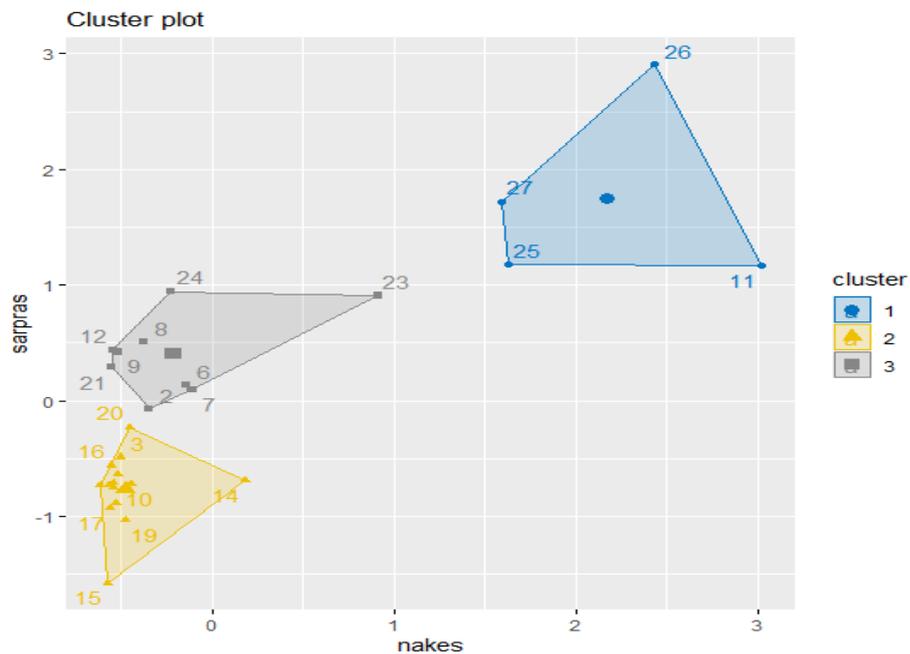
Berdasarkan hasil *clustering* berdasarkan semua variabel, diperoleh bahwa persentasi rasio variansi antar *cluster* dan variansi total adalah 69%. Selanjutnya, dapat dilihat keanggotaan setiap kecamatan terhadap masing-masing *cluster*. Hasil *clustering* berdasarkan semua variabel disajikan pada Tabel 1, dan hasil *cluster-cluster* yang terbentuk secara visual dapat dilihat pada Gambar 3.



Gambar 3. Cluster berdasarkan semua variabel

Pengelompokan kecamatan-kecamatan berdasarkan hanya variabel jumlah tenaga kesehatan (NAKES) dan variabel jumlah sarana dan prasarana (SARPRAS) yang dibagi dengan jumlah penduduk masing-masing kecamatan perlu dipertimbangkan. Hal ini tentunya berkaitan dengan layanan yang dapat diberikan dari kecamatan yang bersangkutan terhadap penduduknya. Oleh karena itu, pada analisis berikutnya dilakukan *clustering* menggunakan variabel NAKES dan SARPRAS.

Hasil *clustering* metode *K-means* dan FCM disajikan pada Tabel 2. Sedangkan visualisasi *cluster* yang dihasilkan disajikan pada pada Gambar 4. Gambar 4 ini menggambarkan visualisasi hasil *Clustering* dimana dengan absis dan ordinat adalah skor baku jumlah tenaga kesehatan (NAKES) dan skor baku jumlah sarana dan prasarana kesehatan (SARPRAS).



Gambar 4. Hasil *Clustering* berdasarkan variabel NAKES dan SARPRAS

Berdasarkan nilai persentase rasio variansi antar *cluster* dengan variansi total, maka dapat disimpulkan bahwa hasil *clustering* berdasarkan hanya variabel NAKES dan SARPRAS yang terkoreksi oleh jumlah penduduk relatif lebih baik dibandingkan dengan hasil *clustering* berdasarkan semua variabel.

Tabel 2. *Cluster* berdasarkan variabel NAKES dan SARPRAS

no	Kecamatan	1	2	3	1	2	3
1	Lumbir	0,001	0,982	0,017	0	1	0
2	Wangon	0,014	0,254	0,732	0	0	1
3	Jatilawang	0,006	0,899	0,095	0	1	0
4	Rawalo	0,000	0,998	0,002	0	1	0
5	Kebasen	0,000	0,994	0,005	0	1	0
6	Kemranjen	0,010	0,074	0,916	0	0	1
7	Sumpiuh	0,013	0,107	0,880	0	0	1
8	Tambak	0,004	0,018	0,978	0	0	1

9	Somagede	0,005	0,031	0,964	0	0	1
10	Kalibagor	0,001	0,989	0,010	0	1	0
11	Banyumas	0,852	0,063	0,085	1	0	0
12	Patikraja	0,007	0,040	0,954	0	0	1
13	Purwojati	0,000	0,995	0,004	0	1	0
14	Ajibarang	0,036	0,712	0,252	0	1	0
15	Gumelar	0,032	0,821	0,147	0	1	0
16	Pekuncen	0,004	0,947	0,049	0	1	0
17	Cilongok	0,002	0,980	0,018	0	1	0
18	Karanglewas	0,001	0,986	0,013	0	1	0
19	Kedungbanteng	0,005	0,960	0,035	0	1	0
20	Baturraden	0,016	0,543	0,442	0	1	0
21	Sumbang	0,006	0,049	0,945	0	0	1
22	Kembaran	0,000	0,996	0,004	0	1	0
23	Sokaraja	0,405	0,165	0,430	0	0	1
24	Purwokerto Selatan	0,053	0,103	0,845	0	0	1
25	Purwokerto Barat	0,865	0,048	0,088	1	0	0
26	Purwokerto Timur	0,841	0,062	0,097	1	0	0
27	Purwokerto Utara	0,941	0,021	0,039	1	0	0
Persentase variansi antar <i>cluster</i> dan total		85%					

Hasil *clustering* berdasarkan variabel jumlah tenaga kesehatan (NAKES) dan jumlah sarana dan prasarana (SARPRAS) adalah sebagai berikut.

- *Cluster 1* terdiri dari kecamatan Banyumas, Purwokerto Barat, Purwokerto Timur, Purwokerto Utara. Pada *cluster 1*, skor baku jumlah NAKES dan jumlah SARPRAS di atas 1,5 dan di atas 1. Hal ini mengindikasikan bahwa jumlah NAKES dan SARPRAS di kecamatan-kecamatan tersebut cukup baik.
- *Cluster 2* yang terdiri dari Kecamatan Lumbir, Jatilawang, Rawalo,...., Kembaran (lihat Tabel 2). Pada *cluster 2* ini skor baku jumlah NAKES dan SARPRAS yang di bawah rata-rata, dan jumlah NAKES dan SARPRASnya paling minimum.
- *Cluster 3* terdiri dari Kecamatan Wangon, Kemranjen, Sumpiuh, Tambak, Somagede, Patikraja, Sumbang, Sokaraja, Purwokerto Selatan. Pada *cluster 3* ini jumlah skor baku jumlah NAKES dan SARPRAS sekitar diantara 0 sampai 1. Jadi masih di atas rata-rata.

4. KESIMPULAN DAN SARAN

Analisis *cluster* dengan metode FCM dan K-mean menghasilkan keanggotaan *cluster* yang sama. Representasi grafik dalam dua dimensi sangat bermanfaat untuk menangkap informasi setiap anggota *cluster*.

Cluster-cluster kecamatan dengan menggunakan jumlah penduduk, NAKES, dan SARPRAS menghasilkan nilai rasio variansi antar kelompok dan total sebesar 69%. *Cluster-cluster* hasil *cclustering* berdasarkan jumlah NAKES dan jumlah SARPRAS menghasilkan nilai rasio variansi antar kelompok dan total sebesar 85%. Sehingga hasil *clustering* berdasarkan variabel jumlah NAKES dan jumlah SARPRAS lebih baik.

Cluster 1 yang terdiri dari kecamatan Banyumas, Purwokerto Barat, Purwokerto Timur, Purwokerto Utara mempunyai jumlah NAKES dan jumlah SARPRAS yang baik. *Cluster 3* terdiri dari Kecamatan Wangon, Kemranjen, Sumpiuh, Tambak, Somagede, Patikraja, Sumbang, Sokaraja, Purwokerto Selatan mempunyai jumlah NAKES dan jumlah SARPRAS yang sedang. *Cluster 2* yang terdiri dari kecamatan-kecamatan diluras cluster 1 dan 3 adalah kecamatan-kecamatan yang mempunyai jumlah NAKES dan SARPRAS yang kurang.

DAFTAR PUSTAKA

- Abonyi, J., Feil, B., *Cluster analysis for data mining and system identification*. Springer Science & Business Media, 2007.
- Badan Pusat Statistik, *Kabupaten Banyumas dalam Angka Tahun 2020*, Badan Pusat Statistik, Banyumas, 2020.
- Khanmohammadi, S., Adibeig, N., & Shanehbandy, S., *An Improved Overlapping K-Means Clustering Method for Medical Applications*, *Expert Systems with Applications*, **67** (2017), 12-18.
- Pusparini, I., *Pengertian Penduduk* [Online], 2012, <https://id.scribd.com/doc/91037202/Pengertian-Penduduk>, diakses pada 8 Oktober 2021.
- Sivanandam, S. N., *Introduction to Fuzzy Logic Using Matlab*, Springer, Berlin, 2007.
- Tokushige, S., Yadohisa, H., Inada, K., *Crisp and Fuzzy K-Means Clustering Algorithms for Multivariate Functional Data*, *Computational Statistics*, **22**(1) (2007), 1-16.

