

**ALGORITMA FUZZY C-MEANS UNTUK MENGELOMPOKKAN
PROVINSI DI INDONESIA BERDASARKAN JUMLAH DESA YANG
MEMILIKI PRASARANA KESEHATAN**
*FUZZY C-MEANS ALGORITHM FOR PROVINCE GROUPING IN
INDONESIA BASED ON THE NUMBER OF VILLAGES THAT HAVE A
HEALTH INFRASTRUCTURE*

Abdul Wahit¹ Hardian Oktavianto² Qurrota a'yun³

¹Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember
email: wahidjunior96@gmail.com

²Dosen Fakultas Teknik, Universitas Muhammadiyah Jember
email: Hardian@unmuhjember.ac.id

³Dosen Fakultas Teknik, Universitas Muhammadiyah Jember
email: Qurrota.ayun@unmuhjember.ac.id

Abstrak

Perbedaan ketersediaan fasilitas kesehatan yang tersebar di Indonesia yang berbeda dan tidak homogen tentu akan menyulitkan bagi pemerintah dalam melakukan evaluasi fasilitas kesehatan tersebut. Oleh karena itu dibutuhkan proses klasifikasi yang disesuaikan dengan kondisi yang ada di masyarakat. Maka dari itu, dibutuhkan mengelompokkan provinsi Indonesia berdasarkan jumlah desa yang memiliki prasarana kesehatan dengan algoritma *Fuzzy C-Means*. Data yang digunakan yaitu jumlah desa yang memiliki prasarana kesehatan di 34 provinsi Indonesia tahun 2018. Dalam menentukan *cluster* optimum, menggunakan validasi metode *Davies Boulding Index*. Dari serangkaian pengujian mulai dari 2 *cluster* sampai 10 *cluster*, dihasilkan *cluster* optimum berada pada 2 *cluster* berdasarkan validasi *Davies Boulding Index* dengan nilai terkecil yaitu 0,387. Pada *cluster* 1 terdiri 30 anggota provinsi dan *cluster* 2 terdiri dari 4 anggota provinsi. Berdasarkan karakteristik data tahun 2018, *cluster* 1 lebih rendah dibandingkan *cluster* 2.

Kata kunci: Prasarana Kesehatan, *Clustering*, *Fuzzy C-Means*, *Davies Boulding Index*.

Abstract

Different opinions about well-known health facilities in Indonesia are different and not homogeneous will certainly make it difficult for the government to do so evaluation of the health facility. Therefore a classification process is needed which is adapted to the existing conditions in the community. Therefore, it is needed grouping Indonesian provinces based on the number of villages that have health infrastructure with the Fuzzy C-Means algorithm. The data used are number of villages that have health infrastructure in 34 provinces of Indonesia years 2018. In determining the optimum cluster, using the Davies method validation Boulding Index. From a series of tests ranging from 2 clusters to 10 clusters, The result is that the optimum cluster is in 2 clusters based on Davies validation The Boulding Index with the smallest value is 0.387. Cluster 1 consists of 30 members province and cluster 2 consists of 4 provincial members. Based on data characteristics in 2018, cluster 1 was lower than cluster 2.

Keywords: *Health Infrastructure, Clustering, Fuzzy C-Means, Davied Boulding index.*

1. PENDAHULUAN

Latar Belakang

Jumlah penduduk Indonesia yang tersebar di 34 provinsi saat ini mencapai 267,7 juta jiwa, berdasarkan data badan pusat statistik

nasional pada tahun 2020 yang tentunya membutuhkan pelayanan kesehatan yang memadai, pelayanan kesehatan dapat dilihat dari ketersediaan prasarana pelayanan yang tersedia di masing-masing wilayah mulai dari desa, kabupaten hingga provinsi. Data prasarana

pelayanan tersebut telah tersedia pada situs resmi badan pusat statistik nasional.

Pada penelitian ini terdapat 6 atribut yaitu, rumah Sakit, rumah sakit bersalin, poliklinik, puskesmas, puskesmas pembantu, dan apotek. Selain menggunakan *Fuzzy C-Means* penulis juga menggunakan validasi metode *Davies Boulding Index*. Validasi ini digunakan untuk menghasilkan informasi dalam menentukan jumlah *Cluster* terbaik.

2. TINJAUAN PUSTAKA

a. Penelitian Terdahulu

Penelitian ini menggunakan beberapa referensi yang berhubungan dengan objek penelitian sebelumnya, diantaranya sebagai berikut.

Hanrizal (2017). Pemanfaatan Algoritma *Clustering* dalam pengelompokan jumlah desa/kelurahan yang memiliki sarana kesehatan menurut provinsi dengan *K-Means*. Data yang digunakan pada penelitian ini adalah data populasi dari tahun 2008 hingga tahun 2014. Hasil dari penelitian ini dapat menjadi masukan bagi pemerintah, provinsi yang menjadi perhatian lebih pada desa/kelurahan yang memiliki sarana kesehatan.

Dwi Retno Sekar Mayangsari, dkk (2019). Pengelompokan jumlah desa yang memiliki sarana kesehatan menurut dengan menggunakan *K-Means Cluster*. Hasil yang didapatkan dari penelitian ini belum merujuk pada fungsionalitas untuk apa data tersebut selanjutnya.

Ni Putu Merliana (2015). Analisa penentuan jumlah *Cluster* terbaik pada metode *K-Means*. Hasil dari penelitian ini akan dijadikan dasar untuk penentuan jumlah *Cluster* dalam melakukan proses *Clustering* dengan metode *K-Means*.

b. Data Mining

Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Hermawati, 2013). *Data mining* adalah analisis dari peninjauan kumpulan data untuk menemukan

hubungan tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dengan bermanfaat bagi pemilik data (Abdurrahman, 2016).

Data mining adalah suatu metode pengolahan data yang mencari pola yang tersembunyi dari sebuah data. Hasil dari pengolahan data dengan menggunakan metode *data mining* dapat digunakan sebagai dasar pengambilan keputusan dimasa depan. *Data mining* memiliki istilah *pattern recognition* atau pengenalan pola dalam pengolahan suatu data. (Ong, 2013).

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam dataset yang berukuran besar. Keluaran dari *data mining* ini akan dijadikan sebagai salah satu dasar pertimbangan pengambilan keputusan di masa depan sehingga *data mining* juga disebut sebagai *knowledge discovery in database* (KKD).

Dari beberapa pengertian di atas, dapat disimpulkan bahwa *data mining* merupakan serangkaian proses dalam pencarian pola, hubungan, penggalian nilai tambah dari data dan informasi yang berukuran besar berupa pengetahuan dengan tujuan menemukan hubungan dan menyederhanakan data agar diperoleh informasi yang dapat dipahami dan bermanfaat dengan bantuan ilmu statistik dan matematika

c. Clustering

Menurut Deka (2014) *Clustering* merupakan salah satu teknik data mining yang digunakan untuk mendapatkan kelompok-kelompok dari objek-objek yang mempunyai karakteristik yang umum di data yang cukup besar. Tujuan utama dari metode *Clustering* adalah pengelompokan sejumlah data atau objek kedalam *Cluster* atau grup sehingga dalam setiap *Cluster* akan berisi data yang semirip mungkin. *Clustering* melakukan pengelompokan data yang didasarkan pada kesamaan antar objek, oleh karena itu klasterisasi digolongkan sebagai metode *unsupervised learning*.

d. Fuzzy C- Means

Ada beberapa algoritma *Clustering* data, salah satu diantaranya adalah *Fuzzy C-Means* (FCM). *Fuzzy C-Means* (FCM) adalah salah satu teknik pengklasteran data keberadaan tiap-tiap titik data dalam suatu *Cluster* ditentukan oleh derajat keanggotaan. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. (Kusumadewi, 2010).

Konsep dasar FCM, pertama kali adalah menentukan pusat *Cluster* yang akan menandai lokasi rata-rata untuk tiap-tiap *Cluster*. Pada kondisi awal, pusat *Cluster* ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan untuk tiap-tiap *Cluster*.

Cara memperbaiki pusat *Cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan dapat dilihat bahwa pusat *Cluster* akan bergerak menuju lokasi yang tepat. Perulangan ini di dasarkan pada minimisasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *Cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. *Output* dari FCM bukan merupakan *fuzzy inference system*, namun merupakan deretan pusat *Cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data. Informasi ini dapat digunakan untuk membangun suatu *fuzzy inference system*. (Kusumadewi, 2010).

Algoritma *Fuzzy C-Means* adalah sebagai berikut:

1. Input data yang akan di *Cluster* X, berupa matriks berukuran n x m (n= jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke-i (i=1,2,...,n), atribut ke-j (j=1,2,...,m).
2. Langkah selanjutnya adalah menentukan beberapa *input* yang dibutuhkan dalam perhitungan *fuzzy c-means*, yaitu:
 - 1) Jumlah *Cluster* (c) ialah banyaknya *Cluster* yang akan dibentuk sesuai dengan kebutuhan pengklasteran.
 - 2) Pangkat (w) ialah nilai eksponen. Maksimum iterasi (MaxIter) merupakan batas pengulangan atau *looping*. *Looping* akan berhenti jika nilai maksimal iterasi sudah tercapai. *Error* terkecil (ξ) berupa batasan nilai yang membuat perulangan

akan berakhir setelah didapatkan nilai *error* yang diharapkan.

- 3) Fungsi objektif awal (P0=0) ialah Suatu fungsi yang akan dioptimumkan (maksimum atau minimum), nilai 0 berarti untuk mendapatkan nilai minimum.
- 4) Iterasi awal (t = 1), iterasi adalah sifat tertentu dari algoritma atau program komputer di mana suatu urutan atau lebih dari langkah algoritma dilakukan secara berulang. Iterasi awal ialah pada perulangan keberapakah program akan dimulai.
3. Membangkitkan bilangan *random* μ_{ik} , i=1,2,...,n; k=1,2,...,c; sebagai elemen-elemen matrik partisi awal i. Hitung jumlah setiap kolom:

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

Q_i ialah jumlah setiap kolom dari nilai *random* sebuah matrik, jumlah Q tergantung dari berapa jumlah kriteria penilaian.

4. Hitung pusat *Cluster* ke-i: V_{kj} , dengan k=1,2,...,c; dan j=1,2,...,m

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

I = iterasi
 μ_{ik} = perubahan matriks
 X_{ij} = atribut

V_{kj} = titik pusat tiap *Cluster*, jumlah V_{kj} tergantung dari berapa *Cluster* yang akan dibentuk dan n ialah jumlah proposal

5. Menghitung fungsi objektif pada iterasi ke-t, P_t

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_j^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right)$$

t merupakan iterasi yang dihitung, jika iterasi dimulai dari 1 maka pada awal perhitungan nilai t ialah 1. Iterasi akan berulang sesuai dengan ketentuan iterasi yang sedang berjalan. Hitung perubahan matrik partisi.

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

Iterasi akan tetap berulang jika nilai atau kondisi-kondisi tertentu belum tercapai, adapun kondisi tersebut ialah jika: $(|Pt - Pt-1| < \xi)$ atau $(t > \text{MaxIter})$ maka berhenti yang mana Pt ialah pusat *Cluster* iterasi ke t kurang dari nilai *error* yang diharapkan atau jika (jumlah iterasi) sudah lebih besar iterasi maksimum. Namun jika iterasi akan diulang lagi dengan 1 akan mengulang proses yang ke-4 atau menghitung pusat *Cluster* lagi. (Kusumadewi, 2010).

e. Metode Davies Boulding Index

Davies Boulding Index (DBI) merupakan salah satu metode yang digunakan untuk mengukur validitas *Cluster* pada suatu metode *Clustering*. Dalam penelitian ini DBI digunakan untuk melakukan validasi data pada setiap *Cluster*. Adapun langkah-langkah untuk menghitung Davies Boulding Index (Sujacka, 2019), yaitu:

1. Menghitung *Sum of Square Within Cluster* (SSW). *Sum of Square Within Cluster* (SSW) merupakan persamaan yang digunakan untuk mengetahui matriks kohesi dalam sebuah *Cluster* ke- i yang dirumuskan sebagai berikut:

$$SSWi = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j c_i)$$

Dengan:

m_i = jumlah data dalam *Cluster* ke- i

c_i = centroid *Cluster* ke- i

$d(x_j c_j)$ = merupakan jarak setiap data terhadap *centroid*

2. Menghitung *Sum of Square Between Cluster* (SSB). *Sum of Square Between Cluster* (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *Cluster* yang dihitung menggunakan persamaan sebagai berikut:

$$SSB_{i,j} = d(c_i, c_j)$$

3. Menghitung jumlah *ratio* (rasio).

Setelah nilai SSW dan SSB diperoleh, kemudian melakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara *Cluster* ke- i dan *Cluster* ke- j . Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{ij} = \frac{SSWi + SSWj}{SSBij}$$

4. Menghitung nilai *Davies Bouldin Index* (DBI).

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *Davies Bouldin Indeks* (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij})$$

Dengan:

k = jumlah *Cluster* yang ditentukan.

Skema *Clustering* yang optimal menurut indeks pengukuran *Davies Bouldin* adalah *Cluster* yang memiliki nilai indeks terkecil atau minimal (Salazar, dkk. 2002).

f. Rapid Miner

Menurut dennis, dkk. (2013). “perangkat lunak yang bersifat terbuka. Rapid miner merupakan sebuah solusi untuk melakukan analisis terhadap data mining, text dan analisis prediksi rapid miner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat menggunakan keputusan yang paling baik”.

Rapid miner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis GUI akan menghasilkan file XML (Extensible Markup Language) yang didefinisikan proses analitis keinginan pengguna untuk ditetapkan ke data. File ini kemudian dibaca oleh rapid miner untuk menjalankan analisis secara otomatis.

g. R-Studio

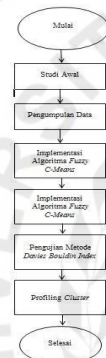
R-Studio adalah aplikasi yang digunakan untuk mempermudah pengolahan bahasa R. Bahasa R yaitu bahasa pemrograman untuk komputasi statistika dan grafik. R-Studio di dirikan oleh JJ Allaire pada tahun 2008. R-

Studio memiliki dua versi, yaitu open source (gratis) dan commercial edition (berbayar). R-Studio juga tidak hanya terbatas dalam bentuk aplikasi dekstop (Windows, MacOS, dan linux), melainkan terdapat versi R-Studio server, yaitu R-Studio yang dapat diakses melalui browser yang terhubung dengan suatu jaringan komputer.

3. METODOLOGI PENELITIAN

a. Rancangan Penelitian

Jenis penelitian yang dilakukan dalam penelitian ini adalah penelitian kuantitatif, yang analisisnya data-data *numerik* (angka). Pada metodologi penelitian terdapat langkah-langkah prosedur penelitian, prosedur tersebut dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian
 Sumber: Diagram Alur Penelitian

b. Studi Literatur

Studi literatur merupakan langkah dalam proses pengumpulan semua informasi yang dibutuhkan untuk membangun sebuah sistem. Informasi ini akan dikumpulkan dari berbagai sumber. Membaca literatur yang ada di jurnal, artikel, buku, dan skripsi adalah salah satu caranya. Persamaan atau perbedaan terhadap penelitian yang dilakukan oleh penulis maupun peneliti lainnya akan diketahui oleh penulis maupun peneliti lainnya yang akan diketa oleh penulis maupun peneliti lainnya.

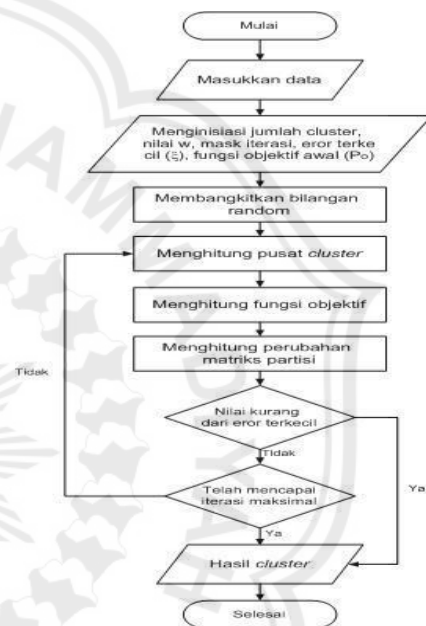
c. DataSet

Penelitian ini mengambil data dari Badan Pusat Statika yang berjudul jumlah desa yang memiliki prasarana kesehatan dari 34 provinsi di Indonesia yang diunduh dari "<https://www.bps.go.id/dynamicTabel/201>

5/09/19/935/jumlah-desakelurahan-yang-memiliki-sarana-kesehatan-menurut-provinsi-2008-2014.html." Data yang digunakan dalam penelitian ini adalah data jumlah desa yang memiliki prasarana kesehatan di Provinsi Indonesia pada tahun 2018.

d. Proses Clustering

Diagram alur algoritma *Fuzzy C-Means* yang digunakan berdasarkan data jumlah desa yang memiliki prasarana kesehatan di provinsi Indonesia, pada umumnya *flowchart Fuzzy C-Means* adalah sebagai berikut pada Gambar 2



Gambar 2. Flowchart Fuzzy C-Means
 Sumber : Flowchart Fuzzy C-Means

e. Metode Davies Boulding Index

Dalam penelitian ini menggunakan teknik pengukuran *Cluster* optimal dengan menggunakan *Davies Bouldin Index*. Adapun tahapan dalam mencari nilai *Davies Bouldin Index* sebagai berikut:

1. Hasil pengelompokkan *Fuzzy C-Means* dan pusat *Cluster* pada iterasi pertama dengan nilai *k* adalah dua.
2. Selanjutnya menghitung *Sum of Square Within Cluster* (SSW) dengan menggunakan rumus.

1. $SSW_i = \frac{74.049+174.325+87.358+150.936+211.220+175.534+444.442+597.408+1004.083+420.139+666.078}{11}$
2. $= \frac{3900,154}{11}$
3. $= 354,559$
4. $SSW_j = \frac{1036,345}{4}$
5. $= 259,086$

3. Kemudian menghitung nilai *Sum of Square Between Cluster* (SSB) dengan menggunakan rumus (2.4).

$$d(C_1, C_2) = \sqrt{\frac{(24.53375079 - 42.1959612)^2 + (37.7996622 - 69.06500576)^2 + (49.88100752 - 121.0946533)^2 + (179.5652652 - 348.493138)^2 + (439.5672474 - 1119.332772)^2 + (112.5 - 202.91)^2}{4}} = 710.741$$

4. Setelah nilai SSW dan SSB diperoleh kemudian melakukan perhitungan pengukuran rasio dengan menggunakan rumus.

$$R_{1,2} = \frac{0,625 + 0,625}{710.741} = 0,625$$

Tabel 1. Nilai *Davies Bouldin Indeks* dengan 2 Cluster

RATIO	c1	c2	MAX	DBI
c1	0.000	0.625	0.625	0.625
c2	0.625	0.000	0.625	

Sumber : Hasil perhitungan excel

Nilai yang diperoleh pada setiap Cluster menunjukkan nilai *Davies Bouldin Index* dengan nilai terkecil sehingga menghasilkan nilai DBI pada 2 Cluster sebesar 0,692.

Skenario Pengujian 2 sampai 10 Cluster menggunakan metode *davies boulding index* untuk mengetahui Cluster optimum yang terdapat di masing- masing Clusternya.

4. HASIL DAN PEMBAHSAN

a. Data Pengujian

Berdasarkan hasil penelitian yang telah didapat dari perhitungan yang telah dilakukan, data jumlah desa yang memiliki prasarana

kesehatan yang telah dikumpulkan di Cluster menggunakan algoritma *Fuzzy C-Means* kemudian diolah untuk mendapatkan hasil Cluster optimum atau Cluster terbaik dengan menggunakan metode *Davis Building Index* Data yang digunakan adalah data jumlah desa yang memiliki prasarana kesehatan pada tahun 2018.

b. Praproses Data

Data diolah menggunakan RStudio yang di cluster menggunakan algoritma *k-means* dari 2 cluster sampai dengan 10 cluster. Output yang dapat dihasilkan dari eksekusi perintah pada RStudio merupakan jumlah iterasi, pusat cluster, fungsi objektif, dan derajat keanggotaan setiap objek terhadap tiap cluster.

c. Penentuan Jumlah Cluster Terbaik

Setelah melalui proses cluster dengan menggunakan algoritma *Fuzzy C-means*, kemudian dilakukan proses dengan menggunakan metode DBI untuk penentuan cluster terbaik. Berikut adalah hasil dari metode DBI.

Tabel 2. Hasil Nilai *Davies Boulding Index*

Cluster	Nilai DBI	Cluster yang dihasilkan	Hasil Anggota Cluster
2	0.383	2	Cluster 1: 30 dan Cluster 2 : 4
3	0.6578	3	Cluster 1: 10, Cluster 2: 21 dan Cluster 3 : 3
4	0.5432	4	Cluster 1: 9, Cluster 2: 10, Cluster 3: 3 dan Cluster 4 :12
5	0.855	5	Cluster 1: 12, Cluster 2: 8, Cluster 3: 9, Cluster 4: 2 dan Cluster 5 : 3
6	0.7218	6	Cluster 1: 4, Cluster 2: 10, Cluster 3: 7, Cluster 4: 8, Cluster 5: 2 dan Cluster 6 : 3
7	0.7998	7	Cluster 1: 5, Cluster 2: 3, Cluster 3: 12,

			Cluster 4: 1, Cluster 5: 8, Cluster 6: 2 dan Cluster 7 : 3	Cluster 1 dan Cluster 2 berdasarkan proses perhitungan yang telah dilakukan sebelumnya. Hasil 2 Cluster yang terdiri dari Cluster 1 dan Cluster 2 sebagai berikut:
8	0.1386	8	Cluster 1: 4, Cluster 2: 5, Cluster 3: 5, Cluster 4: 7, Cluster 5: 2, Cluster 6: 2, Cluster 7: 6 dan Cluster 8 : 3	Pada cluster 1 rentang jumlah bangunan prasarana kesehatan yaitu 10 sampai 199 untuk Rumah Sakit, untuk Rumah sakit pembantu jumlah bangunan prasarana kesehatan yaitu 2 sampai 280, untuk Poliklinik jumlah bangunan prasarana kesehatan yaitu 15 sampai 467, untuk Puskesmas jumlah bangunan prasarana kesehatan yaitu 59 sampai 493, untuk Puskesmas pembantu jumlah bangunan prasarana kesehatan yaitu 3 sampai 1394, dan untuk Apotek jumlah bangunannya 47 sampai 490.
9	0.7292	9	Cluster 1: 9, Cluster 2: 3, Cluster 3: 1, Cluster 4: 3, Cluster 5: 8, Cluster 6: 1, Cluster 7: 2, Cluster 8: 4 dan Cluster 9 : 3	Pada cluster 2 rentang jumlah bangunan prasarana kesehatan yaitu 199 sampai 317 untuk Rumah sakit, untuk Rumah sakit pembantu jumlah bangunan prasarana kesehatan yaitu 441 sampai 1245, untuk Poliklinik jumlah bangunan prasarana kesehatan yaitu 800 sampai 1670, untuk Puskesmas jumlah bangunan prasarana kesehatan yaitu 661 sampai 1170, untuk Puskesmas pembantu jumlah bangunan prasarana kesehatan yaitu 1800 sampai 2338, dan untuk Apotek jumlah bangunannya 740 sampai 2068.
10	11.153	10	Cluster 1: 4, Cluster 2: 5, Cluster 3: 3, Cluster 4: 5, Cluster 5: 1, Cluster 6: 3, Cluster 7: 2, Cluster 8: 2, Cluster 9: 6 dan Cluster 10 : 3	

Sumber : Hasil Perhitungan

Berdasarkan Tabel 4.4 menunjukkan nilai *Davies Boulding Index* pengujian 2 sampai dengan 10 Cluster, maka jumlah Cluster yang paling optimal berdasarkan nilai *Davies Boulding Index* terkecil yaitu 2 Cluster dengan nilai 0,383.

d. Profiling Cluster Optimum

Dari proses klasterisasi menggunakan algoritma *Fuzzy C-Means* dan mencari Cluster terbaik menggunakan metode *Davies Boulding Index*, diketahui Cluster optimum berada pada 2 Cluster. Untuk mengetahui karakteristik dari Cluster optimum pada masing-masing Cluster, maka dilakukan *profiling Cluster* data jumlah desa yang memiliki prasarana kesehatan di setiap provinsi yang masuk ke dalam kelompok

5. KESIMPULAN DAN SARAN

a. Kesimpulan

1. Hasil dari penggunaan algoritma *Fuzzy C-Means* dalam pengelompokan provinsi di Indonesia berdasarkan jumlah desa yang memiliki prasarana kesehatan diperoleh cluster terbaik adalah 2 cluster yang memiliki nilai terendah berdasarkan validasi *Davies Boulding Index* yaitu 0.387 dengan skenario 2 cluster sampai 10 cluster.

2. Hasil cluster optimum menghasilkan informasi dalam menentukan jumlah cluster terbaik, hasil dari pengelompokan yang didapat adalah cluster 1 terdapat 30 provinsi dan cluster 2 terdapat 4 provinsi.

Tabel 3. Anggota cluster 1 dan cluster 2

Anggota Cluster 1	Anggota Cluster 2
Aceh	Sumatera Utara
Sumatera barat	Jawa Barat
Riau	Jawa Tengah
Jambi	Jawa Timur
Sumatera selatan	
Bengkulu	
Lampung	
Kep. bangka belitung	
Kepulauan riau	
Dki jakarta	
D i yogyakarta	
Banten	
Bali	
Nusa tenggara barat	
Nusa tenggara timur	
Kalimantan barat	
Kalimantan tengah	
Kalimantan selatan	
Kalimantan timur	
Kalimantan utara	
Sulawesi utara	
Sulawesi tengah	
Sulawesi selatan	
Sulawesi tenggara	
Gorontalo	
Sulawesi barat	
Maluku	
Maluku utara	
Papua barat	
Papua	

Sumber : Anggota Cluster 1 dan 2

3. Hasil *cluster profiling* menunjukkan tingkat pemerataan prasarana kesehatan pada setiap provinsi di Indonesia pada *cluster 1* memiliki karakteristik data dengan anggota-anggota data yang jumlah prasarana kesehatannya rendah. Pada *cluster 2* memiliki karakteristik data dengan anggota-anggota data yang jumlah prasarana kesehatannya tinggi.

b. Saran

1. Penelitian ini hanya sebagai penerapan algoritma *Fuzzy C- Means*, agar lebih manfaat maka disarankan membangun Sistem Informasi sebagai media implementasi algoritma *Fuzzy C- Means*.

2. Penelitian ini dapat dikembangkan menggunakan metode *Clustering* yang

berbeda, untuk mencari *Cluster* terbaik dapat menggunakan teknik yang berbeda seperti *Silhouette*, *Gap Statistic*, dll

6. DAFTAR PUSTAKA

- Abdurrahman, G., “*Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means*”, Jurnal Sistem & Teknologi Informasi Indonesia, 2016.
- Anonoim., <https://www.bps.go.id/dynamicTable/2015/09/19/935/jumlah-desa-kelurahan-yang-memiliki-sarana-kesehatan-menurut-provinsi-2008-2014.html>. Diakses dan diunduh pada 6 Mei 2020.
- Bwolowalia, P. & Kumar, A. 2014. RBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN. *Inteenatioanal Journal of Computer Application* (0975-8887), IX (105), 17-24.
- Han, J., Kamber, M., Pei, J.: *Data Mining Concept and Techniques, 3rd ed. Morgan Kaufmann-Elsevier, Amsterdam* (2012).
- Hermawati, F. A., 2013, “*Data Mining*”, Penerbit ANDI, Yogyakarta.
- J. O. Ong, “*Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*,” *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10 - 20, 2013.
- Kusumadewi, Sri dan Purnomo Hari. 2010, “*Aplikasi Logika Fuzzy*”, Cetakan Pertama, Graham Ilmu, Yogyakarta.
- M. A. Syakur, B. K. Khotimah, E. M.S. Rochman, and B. D. Satoto, “*Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster*,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018.
- Marliana, N., P., E. Ernawati. Alb. Santoso, J. “*Analisa penentuan jumlah Cluster terbaik pada metode K-Means Clustering*”, *Prosiding Seminar Nasional UNISBANK- ISBN : 978-979-3649-81-8*, 2015.
- Mayangsari, D., R., S. Solikhun. Irawan. “*Pengelompokan Jumlah desa/kelurahan yang memiliki sarana*

- kesehatan menurut Provinsi dengan menggunakan metode K-Means Cluster*
”Jurnal Komik Vol 3. No 1. 2019.
- Paulanda, Z. 2012. *Model Profil Mahasiswa Yang Potensial Drop Out Menggunakan Teknik K-Means Clustering Dan Decision Tree*. Tesis. Universitas Sumatera Utara.
- Rohayani, H., “Analisis Sistem Pendukung Keputusan Dalam Memilih Program Studi Menggunakan Metode Logika Fuzzy”, *Jurnal Sistem Informasi (JSI)*, VOL. 5, NO. 1, April 2013.
- Sadewo. M. G, Windarto. A. P., Andani. S. T., Handrizal. “Pemanfaatan Algoritma Clustering Dalam Mengelompokkan Jumlah Desa/Kelurahan Yang Memiliki Sarana Kesehatan Menurut Provinsi Dengan K-Means”, *Mahasiswa Sistem Informasi STIKOM Tunas Bangsa, Pematiasantar, vol 1, no. 1, Oktober 2017*.
- Sadewo. M. G., Windarto. A. P., dan Hartama. D., “Penerapan Data Mining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering,” *Jurnal Infotekjar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, vol. 2, no. 1, pp. 60-67, 2017.
- Saputra & Rikosakomara. 2018. “Implementasi Fuzzy C-Means dan Model RFM untuk Segmentasi Pelanggan (Studi Kasus: PT. XYZ)”. Surabaya: *Jurnal Teknik ITS* Vol 7, No. 1 (2018) 2337-3520 (2301-928X Print).