

PENGELOMPOKAN KABUPATEN/KOTA DI SUMATERA BARAT BERDASARKAN FAKTOR TERKAIT KEJADIAN DEMAM BERDARAH *DENGUE* DENGAN METODE *FUZZY SUBTRACTIVE* *CLUSTERING*

SRI DELIMAWATI, HAZMIRA YOZZA*, MAIYASTRI

*Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : sridelimawati2121@gmail.com, hazmirayozza@sci.unand.ac.id,
maiyastris@sci.unand.ac.id*

Diterima 15 Desember 2020 Direvisi 29 Desember 2020 Dipublikasikan 12 Januari 2021

Abstrak. Demam berdarah *dengue*(DBD) adalah penyakit infeksi virus akut yang disebabkan oleh virus *dengue*. Provinsi Sumatera Barat merupakan salah satu provinsi di Indonesia yang memiliki angka penderita DBD cukup tinggi. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Sumatera Barat berdasarkan faktor terkait kejadian DBD yakni kejadian banjir, penampungan air, fasilitas dan tenaga kesehatan dan penduduk miskin. Metode yang digunakan dalam penelitian ini adalah metode *fuzzy subtractive clustering*(FSC). Berdasarkan pengolahan dengan metode FSC didapat hasil pengklasteran dengan 2 kluster , 4 kluster, 5 kluster, 7 kluster dan 8 kluster. Dari validitas kluster ditemukan jumlah kluster terbaik yaitu 7 kluster. Berdasarkan karakteristik kluster, secara umum kluster 4 merupakan kluster terbaik dibandingkan dengan kluster lainnya. Kluster 3, kluster 5 dan kluster 7 merupakan kluster terendah, sehingga pada kluster tersebut banyak indikator yang harus mendapat penanganan lebih supaya angka penderita DBD dapat berkurang.

Kata Kunci: Demam Berdarah *Dengue*(DBD), *Fuzzy Subtractive Clustering*(FSC)

1. Pendahuluan

Penyakit demam berdarah *dengue*(DBD) adalah penyakit infeksi virus akut yang disebabkan oleh virus *dengue* dan terutama menyerang anak-anak dengan ciri-ciri demam tinggi mendadak dengan manifestasi perdarahan dan bertendensi menimbulkan shock dan kematian. Sarana penularan demam berdarah sendiri berasal dari gigitan nyamuk *aedes aegypti* dan *aedes albopictus* [4]. Penyakit DBD merupakan salah satu masalah kesehatan masyarakat yang penting di Indonesia. Provinsi Sumatera Barat merupakan salah satu provinsi di Indonesia yang memiliki angka

*penulis korespondensi

penderita DBD cukup tinggi. Banyak faktor yang terkait dengan kejadian DBD, seperti pada kondisi lingkungan yaitu lingkungan yang tidak bersih maupun hal-hal lain yang dapat mempengaruhi kejadian kasus DBD. Pada masing-masing daerah, kondisi faktor-faktor tersebut berbeda sehingga banyaknya kejadian DBD di setiap daerah tersebut juga berbeda. Oleh karena itu perlunya penanganan supaya angka penderita penyakit DBD dapat berkurang. Dalam penanganan tersebut hal pertama yang harus dilakukan adalah mengelompokkan kabupaten/kota mana yang menjadi potensi penyakit DBD tertinggi di Provinsi Sumatera Barat. Pengelompokan tersebut dapat dilakukan berdasarkan faktor-faktor yang mempengaruhi DBD.

Berdasarkan latar belakang tersebut, penulis akan mengkaji lebih lanjut dengan mengelompokkan kabupaten/kota di Sumatera Barat untuk studi kasus demam berdarah tahun 2018 dengan metode *fuzzy subtractive clustering* dan menganalisis jumlah kluster terbaik yang dapat terbentuk berdasarkan indeks validitas kluster.

2. Landasan Teori

2.1. Analisis Kluster

Analisis kluster (*cluster analysis*) merupakan suatu teknik statistik multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik yang dimilikinya. Pengelompokan dilakukan berdasarkan pada sifat kemiripan atau sifat ketidakmiripan antar objek. Objek yang berada dalam kelompok yang sama akan lebih mirip dibandingkan dengan objek pada kelompok yang berbeda [2].

2.2. Himpunan Fuzzy

Himpunan *crisp* adalah nilai keanggotaan x dalam suatu himpunan A , yang sering ditulis dengan $\mu_A(x)$, memiliki dua kemungkinan, yaitu: 1, yang berarti bahwa suatu elemen menjadi anggota dalam suatu himpunan; dan 0, yang berarti bahwa suatu elemen tidak menjadi anggota dalam suatu himpunan[3]. Himpunan *fuzzy* didasarkan pada gagasan untuk memperluas jangkauan fungsi keanggotaan pada himpunan *crisp* sedemikian sehingga fungsi tersebut mencakup bilangan real pada interval $[0,1]$ [6].

2.3. Fuzzy Subtractive Clustering

Fuzzy clustering adalah metode pengklasteran berdasarkan derajat keanggotaan yang mencakup himpunan *fuzzy* sebagai dasar pembobotan bagi pengklasteran. Salah satu pendekatan dari *fuzzy clustering* adalah *fuzzy subtractive clustering* (FSC). Metode ini pertama kali diperkenalkan oleh Stephen L. Chiu pada tahun 1994. Menurut [1], pengklasteran dengan FSC dilakukan pada data yang telah dinormalisasi. Berikut rumus normalisasi data.

$$X_{ij}^* = \frac{X_{ij} - X_{minj}}{X_{maxj} - X_{minj}}, i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (2.1)$$

2.3.1. Pusat Kluster

Penentuan pusat kluster adalah langkah awal pada metode FSC. Pada pendekatan ini, beberapa data memiliki potensi tinggi untuk dapat dijadikan pusat kluster. Data ini adalah data dengan jumlah tetangga yang paling banyak. Data yang sudah terpilih sebagai pusat kluster ini kemudian akan dikurangi potensi terhadap data yang lain. Kemudian algoritma memilih data lain yang memiliki jumlah data terbanyak untuk dijadikan pusat kluster yang lain. Hal ini akan dilakukan berulang-ulang hingga semua data dicobakan [3].

Misalkan terdapat n buah data $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$, dengan $\mathbf{X}_i = (X_{1i}, \dots, X_{mi})$, m adalah banyak variabel. Potensi data ke- k dihitung dari

$$D_k = \sum_{i=1}^n e^{-\frac{4\|\mathbf{X}_k - \mathbf{X}_i\|^2}{r_a^2}} \quad (2.2)$$

dengan \mathbf{X}_k adalah vektor data ke- k , \mathbf{X}_i adalah vektor data ke- i , $\|\mathbf{X}_k - \mathbf{X}_i\|$ adalah jarak *euclidean* antara \mathbf{X}_k dengan \mathbf{X}_i , n adalah jumlah data dan r_a adalah jari-jari setiap variabel data.

Misalkan \mathbf{X}_{c1} adalah data yang terpilih sebagai pusat kluster pertama, dengan potensi sebesar D_{c1} . Nilai D_{c1} selanjutnya menjadi potensi kluster pertama. Pada tahap selanjutnya, dihitung kembali potensi data yang lain dengan mengurangi potensi sebelumnya dengan potensi pusat kluster pertama tersebut melalui persamaan

$$D'_k = D_k - D_{c1} * e^{-\frac{4\|\mathbf{X}_k - \mathbf{X}_{c1}\|^2}{r_b^2}} \quad (2.3)$$

dengan D'_k adalah nilai potensi baru data ke- k dan r_b adalah jari-jari setiap variabel data. Biasanya r_b bernilai lebih besar dibandingkan dengan r_a , $r_b = q * r_a$. Nilai q adalah *squash factor* biasa dinyatakan $q = 1.25$. Setelah potensi baru dari semua data diperoleh, data dengan potensi tertinggi dipilih sebagai pusat kluster yang kedua. Selanjutnya didapatkan pusat kluster kedua, nilai potensi tiap-tiap data dikurangi kembali, demikian seterusnya.

Proses penentuan pusat kluster dilakukan secara iteratif. Iterasi tersebut dihentikan berdasarkan dua nilai, yaitu *accept ratio* dan *reject ratio*. *Accept ratio* dan *reject ratio* keduanya merupakan suatu bilangan pecahan yang bernilai 0 sampai 1. *Accept ratio* merupakan batas bawah suatu data yang menjadi calon pusat kluster diterima menjadi pusat kluster, sedangkan *reject ratio* merupakan batas atas dimana suatu data yang menjadi calon pusat kluster tidak diterima untuk menjadi pusat kluster. Pada suatu iterasi, jika ditemukan suatu data dengan potensi tertinggi (misal dengan pusat kluster \mathbf{X}_k dengan potensi D_k), maka dilanjutkan dengan mencari *ratio* potensi data tersebut dengan potensi tertinggi yang diperoleh pada iterasi pertama (misal pusat kluster pertama \mathbf{X}_{c1} dengan potensi data terbesar kluster pertama D_{c1}). Hasil bagi antara D_k dengan D_{c1} ini kemudian disebut dengan *ratio*.

Untuk menentukan pusat kluster terdapat tiga kondisi yang bisa terjadi dalam suatu iterasi yaitu:

- (1) jika $ratio > accept\ ratio$, maka data tersebut diterima sebagai pusat kluster baru,
- (2) jika $reject\ ratio < ratio \leq accept\ ratio$, maka data tersebut diterima sebagai pusat kluster baru hanya jika data tersebut terletak pada jarak yang cukup jauh dengan pusat kluster yang lainnya,
- (3) jika $ratio \leq reject\ ratio$, maka sudah tidak ada lagi data yang dipertimbangkan untuk menjadi calon pusat kluster, iterasi dihentikan.

2.3.2. Derajat Keanggotaan

Berdasarkan data sampel dapat diduga simpangan baku dari masing-masing variabel menggunakan rumus

$$\sigma_j = \frac{r_a * (X_{maxj} - X_{minj})}{\sqrt{8}}, \quad j = 1, 2, \dots, m. \quad (2.4)$$

Pusat kluster (C) yang dihasilkan dari tahap pertama tadi dan simpangan baku (σ) pada persamaan (2.4) digunakan untuk menentukan nilai parameter fungsi keanggotaan *gauss*. Dengan fungsi tersebut, derajat keanggotaan suatu titik data X_i pada kluster ke- l ($l = 1, 2, \dots, p$) adalah [3]:

$$\mu_{li} = e^{-\sum_{j=1}^m \frac{(X_{ij} - C_{lj})^2}{2\sigma_j^2}}. \quad (2.5)$$

2.3.3. Keanggotaan Kluster

Dalam menentukan keanggotaan kluster, digunakan nilai derajat keanggotaan yang telah didapatkan sebelumnya. Derajat keanggotaan terbesar menunjukkan bahwa kecenderungan tertinggi suatu data untuk masuk menjadi anggota kluster tersebut [3].

2.4. Algoritma Fuzzy Subtractive Clustering

Algoritma *fuzzy subtractive clustering* adalah sebagai berikut [3].

- (1) Menginput data yang akan diklusterkan berupa matriks (\mathbf{X}_{ij}), dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, m$.
- (2) Menetapkan nilai jari-jari (r_a), *squash factor* (q), *accept ratio* dan *reject ratio*.
- (3) Menentukan nilai minimum (X_{min}) dan maksimum (X_{max}).
- (4) Menormalisasi setiap data menggunakan persamaan (2.1).
- (5) Menentukan potensi awal tiap titik data, D_k ($k = 1, 2, \dots, n$) dengan persamaan (2.2).
- (6) Menentukan nilai $D_{c1} = D_{kmax}$, khusus untuk iterasi ke-1.
- (7) Menghitung nilai $Ratio = \frac{D_{kmax}}{D_{c1}}$, untuk iterasi ke-2 dan seterusnya. Setelah nilai *ratio* diperoleh, ada tiga keadaan yang dapat terjadi:
 - (a) jika nilai $ratio > accept\ ratio$, calon pusat kluster dapat diterima sebagai pusat kluster baru,

(b) jika nilai $ratio < accept\ ratio$ dan nilai $ratio > reject\ ratio$, calon baru akan diterima jika keberadaannya cukup jauh dari pusat kluster yang telah ada. Misal \mathbf{V} adalah calon pusat kluster. Prosedur yang dilakukan adalah:

i. menghitung jarak calon pusat kluster \mathbf{V} dengan pusat kluster yang lain C_l dengan rumus:

$$Sd_l = \sum_{j=1}^m \left(\frac{V_j - C_{lj}}{r_j} \right)^2 \quad (2.6)$$

ii. menentukan jarak terdekat \mathbf{V} dengan pusat kluster lain Md , hitung

$$Mds = \sqrt{Md};$$

A. jika $(ratio + Mds) \geq 1$, calon pusat kluster diterima sebagai pusat kluster baru,

B. jika $(ratio + Mds) < 1$, calon pusat kluster tidak diterima dan tidak akan dipertimbangkan kembali sebagai pusat kluster dan potensi data tersebut diset menjadi 0,

(c) jika nilai $ratio < accept\ ratio$ dan nilai $ratio < reject\ ratio$, sudah tidak ada lagi calon pusat kluster baru dan iterasi dihentikan.

(8) Mengurangi potensi titik-titik data yang lain dengan persamaan (2.3).

(9) Menentukan nilai D'_{kmax} .

(10) Mengulangi langkah 7-9 sampai proses iterasi berhenti.

(11) Mengembalikan pusat kluster dari bentuk ternormalisasi ke bentuk semula.

$$C_{lj_{denorm}} = C_{lj} * (X_{maxj} - X_{minj}) + X_{minj} \quad (2.7)$$

(12) Menghitung nilai simpangan baku kluster dengan persamaan (2.4).

(13) Menghitung derajat keanggotaan suatu titik data X_i pada kluster ke- l ($l = 1, 2, \dots, p$) dengan persamaan (2.5).

(14) Menentukan keanggotaan kluster.

2.5. Validitas Kluster

Salah satu indeks validitas yang hanya melibatkan nilai keanggotaan adalah indeks *modified partition coefficient* (MPC). MPC merupakan indeks validitas hasil perbaikan dari metode *partition coefficient* (PC). Nilai MPC didefinisikan dengan persamaan:

$$MPC = 1 - \frac{p}{p-1} (1 - PC) \quad (2.8)$$

Adapun persamaan indeks PC yaitu :

$$PC = \frac{1}{n} \sum_{l=1}^p \sum_{i=1}^n (\mu_{li})^2 \quad (2.9)$$

Nilai MPC berkisar antara $0 \leq MPC \leq 1$. Penentuan hasil kluster optimal dari beberapa jumlah kluster yang dihasilkan ditentukan dari nilai MPC yang terbesar[5].

3. Metode Penelitian

Data yang digunakan adalah data sekunder mengenai faktor terkait kejadian DBD kabupaten/kota di Sumatera Barat pada tahun 2018 yang bersumber dari Survei Sosial Ekonomi Nasional(Susenans) Badan Pusat Statistik Provinsi Sumatera Barat tahun 2018.

Variabel yang digunakan dalam penelitian ini adalah persentase penduduk miskin (X_1), jumlah penderita penyakit DBD (X_2), jumlah desa/kelurahan yang mengalami banjir (X_3), jumlah puskesmas (X_4), jumlah dokter umum (X_5) dan persentase rumah tangga dengan penampungan air hujan (X_6).

Tahap-tahap yang dilakukan adalah :

- (1) mendeskripsikan masing-masing variabel yang digunakan dalam penelitian,
- (2) melakukan pengklasteran menggunakan algoritma FSC,
- (3) menentukan rekomendasi klaster terbaik dengan menggunakan indeks validitas klaster,
- (4) menginterpretasikan hasil pengklasteran.

4. Pembahasan

4.1. Hasil Pengklasteran

Proses klastering menggunakan algoritma FSC dengan mengambil $r = 0.40, 0.50, 0.60, 0.70, 0.80, 0.90$. Berdasarkan output menggunakan software SPSS 16, Matlab 2017b maupun perhitungan manual menggunakan Microsoft Excel, banyak klaster yang terbentuk dengan $r = 0.40, 0.50, 0.60, 0.70, 0.80, 0.90$ adalah sebagai berikut.

Tabel 1. Banyaknya Klaster yang Terbentuk dengan $r = 0.40, 0.50, 0.60, 0.70, 0.80, 0.90$

Jari-jari(r)	Banyaknya Klaster
0.40	8
0.50	7
0.60	7
0.70	5
0.80	4
0.90	2

Dari Tabel 1 dapat dilihat bahwa banyaknya klaster yang diperoleh dengan $r = 0.40$ adalah 8 klaster, $r = 0.50, 0.60$ adalah 7 klaster, $r = 0.70$ adalah 5 klaster, $r = 0.80$ adalah 4 klaster dan $r = 0.90$ adalah 2 klaster.

4.2. Rekomendasi Klaster Terbaik

Perhitungan nilai indeks *modified partition coefficient*(MPC) dilakukan pada seluruh jari-jari yang membentuk beberapa jumlah klaster yang telah dianalisis

menggunakan metode FSC. Berikut adalah nilai indeks MPC untuk $r = 0.40, 0.50, 0.60, 0.70, 0.80, 0.90$.

Tabel 2. Nilai MPC

Jari-jari (r)	Banyaknya Klaster	MPC
0.40	8	0.467
0.50	7	0.462
0.60	7	0.558
0.70	5	0.393
0.80	4	0.367
0.90	2	0.382

Berdasarkan Tabel 2 nilai indeks MPC yang terbesar terdapat pada jari-jari (r) = 0.60. Hal ini menunjukkan klaster tersebut merupakan jumlah klaster yang terbaik yang digunakan dalam permasalahan ini. Jari-jari (r) = 0.60 diperoleh kesamaan kecenderungan data yang masuk pada setiap klaster, berikut klaster yang terbentuk dengan $r = 0.60$.

Tabel 3. Hasil Pengklasteran dengan $r = 0.60$

Klaster	Nama Kabupaten/Kota
Klaster 1	Kab. Solok Selatan, Kota Solok, Kota Sawahlunto, Kota Padang Panjang, Kota Bukittinggi, Kota Payakumbuh, Kota Pariaman
Klaster 2	Kab. Sijunjung, Kab. Solok, Kab. Pasaman, Kab. Dharmasraya, Kab. Pasaman Barat
Klaster 3	Kab. Padang Pariaman, Kab. Lima Puluh Kota,
Klaster 4	Kab. Tanah Datar, Kab Agam
Klaster 5	Kabupaten Mentawai
Klaster 6	Kota Padang
Klaster 7	Kab. Pesisir Selatan

4.3. Karakteristik Klaster Terbaik

Setelah klaster terbentuk, diambil rata-rata dari masing-masing faktor yang terkait dengan DBD terhadap seluruh objek penelitian (19 kabupaten/kota di Sumatera Barat), dimana faktor-faktor tersebut yaitu penduduk miskin, penderita DBD, desa/kelurahan yang megalami banjir, puskesmas, dokter umum dan rumah tangga dengan penampungan air hujan (\bar{X}). Selanjutnya untuk masing-masing klaster juga diambil rata-rata pada masing-masing variabel (\bar{X}_l). Pada setiap peubah di dalam klaster diberi tanda, jika $\bar{X}_l > \bar{X}$ maka diberi tanda positif (+), sedangkan jika

$\bar{X}_l < \bar{X}$ maka diberi tanda negatif (-). Adapun karakteristik hasil pengklasteran disajikan pada tabel berikut.

Tabel 4. Karakteristik Hasil Pengklasteran

Klaster	X_1	X_2	X_3	X_4	X_5	X_6
Klaster 1	-	-	-	-	-	-
Klaster 2	+	-	+	+	+	-
Klaster 3	+	-	+	+	-	+
Klaster 4	-	+	-	+	+	-
Klaster 5	+	-	-	-	-	+
Klaster 6	-	+	+	+	+	-
Klaster 7	+	+	+	+	-	-

Pada penelitian ini variabel persentase penduduk miskin (X_1), jumlah penderita DBD (X_2), jumlah desa/kelurahan yang mengalami banjir (X_3) dan persentase rumah tangga dengan penampungan air hujan (X_6) jika bertanda (-) maka kemungkinan terjadinya DBD sedikit pada suatu klaster tersebut. Untuk jumlah puskesmas (X_4) dan jumlah dokter umum (X_5) jika bertanda (+) maka kemungkinan terjadinya DBD sedikit pada suatu klaster tersebut.

Berdasarkan Tabel 4 dapat dilihat bahwa klaster 1, klaster 2 dan klaster 6 dapat dikatakan tingkat penanganannya baik. Klaster 4 dapat dikatakan tingkat penanganannya sangat baik. Klaster 3, klaster 5 anganannya kurang baik.

5. Kesimpulan

Berdasarkan hasil penelitian ini, dapat diambil kesimpulan bahwa pengelompokan berdasarkan faktor terkait kejadian DBD dengan menggunakan metode FSC dimana jumlah klaster terbaiknya adalah 7 klaster dengan jari-jari(r) = 0.60. Berdasarkan karakteristik klaster, secara umum klaster 4 dapat dikatakan klaster terbaik. Klaster 3, klaster 5 dan klaster 7 adalah klaster terendah, sehingga harus mendapat penanganan lebih supaya angka penderita penyakit DBD dapat berkurang.

6. Ucapan Terima kasih

Penulis mengucapkan terimakasih kepada Ibu Izzati Rahmi, H.G., M.Si, Ibu Dr. Arrival Rince Putri dan Ibu Monika Rianti Helmi M.Si yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Chiu, S.L. 1994. Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent and Fuzzy Systems*. Rockwell Science Center. California.
- [2] Johnson R, Wichern D. 2007. *Applied Multivariate Statistical Analysis*, Sixth Edition. New Jersey : Pearson Education.
- [3] Kusumadewi dan Purnomo. 2004. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Edisi 2. Graha Ilmu. Yogyakarta.

- [4] Siregar, Faziah A. 2004. Epidemiologi dan Pemberantasan Demam Berdarah Dengue (DBD) di Indonesia. Fakultas Kesehatan Masyarakat. Universitas Sumatera Utara. <http://repository.usu.ac.id/bitstream/123456789/3673/3/fkmfazidah3.pdf.txt>
- [5] Xie, N., L.Hu, N.Luktarhan, dan K.Zhao. 2011. A Classification of Cluster Validity Indexes Based on Membership Degree and Application. *Web Information System and Mining*. 43 – 50.
- [6] Yan, J., M. Ryan and J. Power. (1994). *Using Fuzzy Logic Towards Intelligent Systems*. Prentice Hall International, London.