

## Analisis Cluster untuk Pengelompokan Desa Berdasarkan Indikator Penyakit Diare

Rahmawati\*<sup>1</sup>, Muliady Faisal<sup>2</sup>

<sup>1,2</sup>Universitas Sulawesi Barat

e-mail: \*<sup>1</sup>rahmah@unsulbar.ac.id, <sup>2</sup>muliadyfaisal@unsulbar.ac.id

### Abstrak

Penelitian ini bertujuan mengelompokkan desa berdasarkan indikator penyakit diare. Metode K-Means adalah salah satu teknik analisis cluster yang dapat dilakukan untuk mengelompokkan daerah yang memiliki kemiripan. Metode penelitian dimulai dengan (1) melakukan observasi penyebab penyakit diare, (2) mengumpulkan data, (3) melakukan pengelompokan daerah menggunakan analisis cluster, kemudian (4) menginterpretasikan hasil analisis. Sebanyak 51 desa dikelompokkan berdasarkan 5 variabel yaitu: Keberadaan tempat buang air di desa setempat, Kondisi pemukiman, Sumber air minum dan potensi banjir. Analisis cluster dengan menggunakan algoritma K-Means menghasilkan Cluster 1 sebanyak 4 desa, Cluster 2 sebanyak 37 desa, Cluster 3 sebanyak 2 desa, dan Cluster 4 sebanyak 8 desa. Hasil pengelompokan memberikan informasi penyebab penyakit diare dan menjadi acuan untuk pencegahannya di setiap cluster.

**Kata kunci:** Penyakit Diare, K-Means, Analisis Cluster

### 1 PENDAHULUAN

Diare merupakan penyakit penyebab kematian tertinggi di dunia, berdasarkan data WHO pada tahun 2016. Penyakit Diare merupakan penyakit endemis dan juga merupakan penyakit potensial KLB (Kejadian Luar Biasa-Keracunan) yang sering disertai dengan kematian. Pada tahun 2015 di Indonesia terjadi 18 kali KLB Diare yang tersebar di 11 provinsi, 18 kabupaten/kota, dengan jumlah penderita 1.213 orang dan kematian 30 orang (Kemenkes RI, 2015).

Kondisi tersebut di atas dialami oleh hampir seluruh negara di dunia, terlebih pada negara berkembang. Tidak terkecuali di setiap daerah di Indonesia, kurangnya pengetahuan dan kesadaran masyarakat menjadi faktor penyebab tingginya penderita penyakit diare. Persoalan ini diidentifikasi dan ditanggulangi mulai dari tingkat daerah di setiap provinsi yang ada di Indonesia, seperti halnya di kabupaten Majene provinsi Sulawesi Barat. Menurut data dari dinas kesehatan setempat, kasus diare menjadi salah satu kasus yang terbanyak dilaporkan pada setiap tahunnya. Misalkan pada tahun 2014 kasus ini dilaporkan penderita penyakit diare sebanyak 4.900 jiwa, angka ini jauh melampaui angka penderita penyakit lainnya seperti Demam Berdarah Dengue (DBD) dan Campak.

Pencegahan wabah penyakit dapat dilakukan secara efektif melalui kerjasama yang baik antara pihak yang berwenang dalam masalah kesehatan masyarakat dan masyarakat itu sendiri. Banyaknya penderita penyakit diare di setiap daerah berbeda, hal ini disebabkan oleh tingkat pengetahuan penduduk terkait penyakit tersebut. Upaya penanggulangannya dilakukan mulai dari pencegahan sampai pada pengobatan dan pemulihan penderita penyakit tersebut.

Urgensi penanggulangan penyakit diare didasarkan pada identifikasi penyebab dan jumlah penderita penyakit tersebut di setiap daerah. Terkait hal tersebut, analisis cluster dapat

digunakan untuk mengelompokkan desa berdasarkan penyakit diare dan persoalannya yang kemudian dapat digunakan sebagai acuan untuk penanggulangannya.

### 1.1 Penyakit Diare

Menurut Departemen Pendidikan Nasional (2015), Epidemio atau wabah adalah penyakit menular yg berjangkit dengan cepat, salah satu diantaranya adalah diare yang merupakan suatu gangguan kesehatan, dimana penderita mengalami muntah-muntah dan diare tak hanya sekali atau dua kali dengan selang waktu yang tidak lama. Sanitasi lingkungan yang tidak terjaga dengan baik juga memudahkan kuman penyebab diare untuk berkembang biak. Hujan yang terus menerus sehingga menimbulkan banjir dan lingkungan yang kotor, sangat potensial menimbulkan wabah diare. Tingkat ancaman epidemio penyakit diare ditentukan berdasarkan beberapa faktor, salah satunya adalah faktor lingkungan, yaitu sarana air bersih dan jamban (Evayanti et al. 2014). Faktor yang lain adalah ketersediaan sarana tempat pembuangan sampah, ketersediaan pembuangan air limbah, dan *personal hygiene* ibu tersebut, Mafazah (2012). Berdasarkan informasi tersebut yang disesuaikan dengan data BPS tahun 2014, faktor paling berpengaruh terhadap epidemio penyakit tertera dalam Tabel 1.

Tabel 1 Faktor penyebab penyakit diare

No	Faktor Penyakit Muntaber	Sumber Data
1	Tempat buang air besar	BPS 2014b
2	Keberadaan pemukiman kumuh	BPS 2014b
3	Jumlah kejadian diare	BPS 2014b
4	Sumber air minum	BPS 2014b
5	Bahaya banjir	BIG 2012

### 1.2 Analisis Cluster

Analisis cluster (*Cluster analysis*) adalah pekerjaan mengelompokkan data (objek) yang didasarkan pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya (Tan, 2006). Tujuannya adalah agar objek-objek yang bergabung dalam sebuah kelompok (cluster) merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam kelompok lain. Lebih besar kemiripannya (homogenitas) dalam kelompok dan lebih besar perbedaannya di antara kelompok yang lain, konsep inilah yang akan dibahas dalam pengelompokan.

Tujuan analisis cluster dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data, sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari prototipe kelompok yang paling representatif terhadap data. (Prasetyo, 2012). Beberapa teknik analisis cluster dibedakan berdasarkan struktur kelompok dan keanggotaan dalam kelompok. Menurut struktur, pengelompokan membagi set data ke dalam sejumlah kelompok yang tidak tumpang-tindih (*overlap*) antara satu kelompok dengan kelompok yang lain. Artinya setiap data hanya menjadi anggota satu kelompok. Metode K-Means dan DBSCAN masuk dalam kategori ini. Menurut keanggotaan, pengelompokan membolehkan sebuah data menjadi anggota di lebih dari satu kelompok, pengelompokan ini dapat dilakukan dengan metode *Fuzzy C-Means*.

### 1.3 Konsep K-Means

*K-Means* merupakan salah satu metode pengelompokan data yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data yang ada ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain.

Algoritma pengelompokan dengan K-Means adalah sebagai berikut:

1. Menentukan jumlah kelompok.
2. Alokasikan data ke dalam kelompok secara acak.
3. Hitung pusat kelompok (sentroid/rata-rata) dari data yang ada di masing-masing kelompok.
4. Alokasikan masing-masing data ke sentroid/rata-rata terdekat.
5. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai sentroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

Lokasi sentroid (rata-rata) fitur ke- $i$  pada langkah 3 di atas menggunakan formula

$$C_i = \frac{1}{M} \sum_{j=1}^M \mathbf{x}_j \quad (1)$$

Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, diantaranya Euclidean (Bezdek, 1981), Manhattan/City Block (Miyamoto dan Agusta, 1995) dan Minkowsky (Miyamoto dan Agusta, 1995).

Pengukuran jarak pada ruang jarak (distance space) Euclidean menggunakan formula:

$$\mathbf{D}(\mathbf{x}_2, \mathbf{x}_1) = \|\mathbf{x}_2 - \mathbf{x}_1\|_2 = \sqrt{\sum_{j=1}^p |\mathbf{x}_{2j} - \mathbf{x}_{1j}|^2} \quad (2)$$

$\mathbf{D}$  adalah jarak antara  $\mathbf{x}_2$  dan  $\mathbf{x}_1$ , dan  $|\circ|$  adalah nilai mutlak.

Pada langkah 4, pengalokasian kembali data ke dalam masing-masing kelompok dalam metode K-Means didasarkan pada perbandingan jarak antara data dengan sentroid setiap kelompok yang ada. Data dialokasikan ulang secara tegas ke kelompok yang mempunyai sentroid dengan jarak terdekat dari data tersebut. Pengalokasian ini dapat dirumuskan sebagai berikut:

$$a_{i1} = \begin{cases} 1 & d = \min\{\mathbf{D}(\mathbf{x}_i, C_1)\} \\ 0 & \text{lainnya} \end{cases} \quad (3)$$

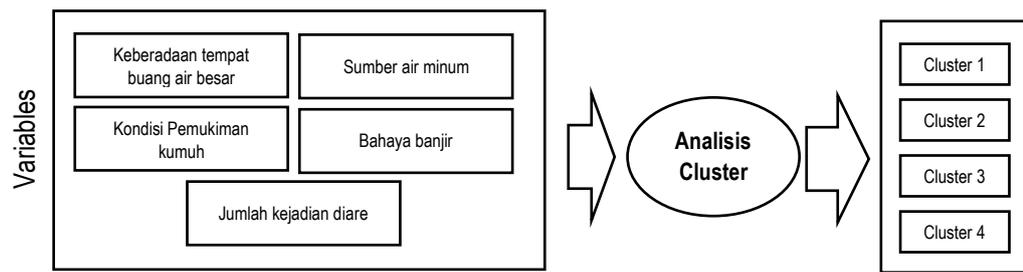
$a_{i1}$  adalah nilai keanggotaan titik  $\mathbf{x}_i$  ke pusat  $C_1$ ,  $d$  adalah jarak terdekat dari data  $\mathbf{x}_i$  ke  $K$  kelompok setelah dibandingkan, dan  $C_1$  adalah sentroid ke-1.

Fungsi objektif yang digunakan untuk K-Means ditentukan berdasarkan jarak dan nilai keanggotaan data dalam kelompok. Fungsi objektif yang digunakan adalah sebagai berikut (MacQueen, 1967):

$$\mathbf{J} = \sum_{i=1}^N \sum_{l=1}^K a_{il} \mathbf{D}(\mathbf{x}_i, C_l)^2 \quad (4)$$

$N$  adalah jumlah data,  $K$  adalah jumlah kelompok,  $a_{il}$  adalah nilai keanggotaan titik  $\mathbf{x}_i$  ke pusat kelompok  $C_l$ ,  $C_l$  adalah pusat kelompok ke- $l$ , dan  $\mathbf{D}(\mathbf{x}_i, C_l)$  adalah jarak  $\mathbf{x}_i$  ke kelompok  $C_l$  yang diikuti.  $a$  mempunyai nilai 0 atau 1. Apabila suatu data merupakan anggota suatu kelompok, nilai  $a_{i1} = 1$ . Jika tidak  $a_{i1} = 0$ .

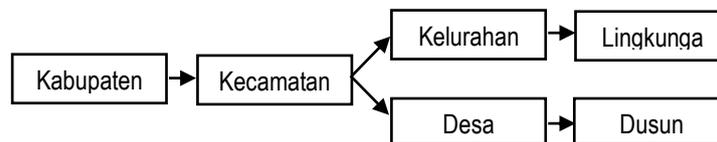
Kerangka konseptual penelitian digambarkan dalam bagan berikut:



Gambar 1 Kerangka Konseptual Penelitian

## 2 METODE PENELITIAN

Pengumpulan data dilakukan melalui pengamatan langsung, wawancara dan pengisian kuesioner. Jumlah penderita penyakit diare diperoleh melalui kerjasama dengan Dinas Kesehatan Majene dan Puskesmas di daerah Kecamatan Banggae yaitu Puskesmas Banggae 1 dan Puskesmas Totoli. Data penderita diare tercatat selama 1 tahun terakhir. Struktur daerah di wilayah Majene diperlihatkan dalam bagan pada Gambar 2



Gambar 2 Bagan klasifikasi wilayah Majene

Desa dalam penelitian ini didefinisikan sebagai elemen terkecil dalam suatu wilayah, dalam hal ini lingkup penelitian adalah Lingkungan dan Dusun. Daerah kecamatan Banggae merupakan obyek penelitian yang secara keseluruhan terdiri dari Lingkungan dan Dusun sebanyak 51 daerah.

Berdasarkan kajian literatur yang dilakukan sebelumnya maka variabel penelitian diantaranya:

- Keberadaan tempat buang air di desa setempat (X1)
- Kondisi pemukiman: kumuh atau bersih (X2)
- Jumlah penderita diare selama 3 tahun terakhir (X3)
- Sumber air minum (X4)
- Potensi banjir (X5)

Hasil pengukuran dinyatakan dalam bentuk persentase keterpenuhan aspek yang dikaji. Beberapa pertanyaan diberikan untuk mengukur satu variabel.

### Implementasi K-Means

Pengelompokan 51 desa berdasarkan indikator penyakit diare dilakukan menggunakan bantuan software IBM SPSS 20 dengan algoritma sebagai berikut:

1. Menentukan banyaknya kelompok.

Dipilih banyaknya kelompok/kluster  $K = 4$

2. Data dibagi secara acak menjadi 4 kelompok, yaitu:

$$\mathbf{x}_1 = \{V1, V2\} = \{(96, 65, 12, 100, 0), (85, 63.5, 8, 100, 0)\}$$

$$\mathbf{x}_2 = \{V3, V4\} = \{(90, 80, 13, 100, 75), (90, 73.5, 7, 100, 75)\}$$

$$\mathbf{x}_3 = \{V5, V6\} = \{(80, 72.5, 1, 100, 75), (75, 73.5, 7, 100, 75)\}$$

$$\mathbf{x}_4 = \{V7, V8, \dots, V51\} = \{(90, 68.5, 7, 100, 75), (85, 64, 0, 100, 100), \dots, (70, 76, 1, 100, 75)\}$$

3. Menghitung sentroid dari data yang ada di masing-masing kelompok.

$$C_1 = (90.5, 64.25, 10, 100, 0)$$

$$C_2 = (90, 76.75, 10, 100, 75)$$

$$C_3 = (77.5, 73, 4, 100, 75)$$

$$C_4 = (107.6, 69.3, 11.8, 100, 72.78)$$

4. Cluster baru didasarkan pada jarak terdekat masing-masing data ke sentroid. Jarak  $V_3$  ke sentroid  $C_1$  dihitung dengan menggunakan jarak Euclidean

$$\begin{aligned} D(V_3, C_1) &= \sqrt{(90 - 90.5)^2 + (80 - 64.25)^2 + (13 - 10)^2 + (100 - 100)^2 + (75 - 0)^2} \\ &= 76.7 \end{aligned}$$

5. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai sentroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

### 3. HASIL DAN PEMBAHASAN

Hasil iterasi menunjukkan bahwa cluster yang terbentuk adalah:

- **Cluster 1** : Pangali-ali, Cilallang, Tanangan, Battayang
- **Cluster 2** : Pa'leo, Timbo-timbo, Pa'leo Towanda, Panggalo, Salabose, Rusung, Parappe, Binanga, Lipu, Tulu, Kampung Baru, Copala, Galung Utara, Galung Barat, Garo'go, Passarang, Passarang Selatan, Moloku, Mangge, Kalasa, Rangas Timur, Rangas Tammalassu, Rangas Pa'besoang, Rangas Barat, Deteng-deteng, Pamboborang, Pamboborang Selatan, Galung Pa'ara, Galung Pa'ara Selatan, Lambe, Kanappe, Labu-labuang, Batu-batu, Alinduang
- **Cluster 3** : Tanjung Batu, Galung Selatan
- **Cluster 4** : Tanangan Barat, Saleppa, Galung Tengah, Camba, Camba Utara, Teppo Barat, Konja, Konja Selatan,

Jarak sentroid masing-masing cluster diperlihatkan dalam Tabel 5.2

Tabel 5.2 Jarak sentroid masing-masing cluster

**Distances between Final Cluster Centers**

Cluster	1	2	3	4
1		76,643	603,981	44,703
2	76,643		614,040	32,186
3	603,981	614,040		610,594
4	44,703	32,186	610,594	

### 4 KESIMPULAN

Analisa lebih lanjut masih perlu dilakukan untuk mengidentifikasi secara detail penanggulangan penyakit diare untuk setiap cluster berdasarkan karakteristik daerahnya.

### DAFTAR PUSTAKA

Dinas Kesehatan Majene. 2014. Profil Kesehatan Tahun 2014. PMD. Majene

- Evayanti, N.K.E, I.N Purna, I.K Aryana. 2014. Faktor-faktor yang berhubungan dengan kejadian Diare pada Balita yang Berobat ke Badan Rumah Sakit Umum Tabana. *Urnal Kesehatan Lingkungan*. 4(2): 134-139
- Mafazah, L. 2012. Ketersediaan Sarana Sanitasi Dasar, Personal Hygiene Ibu dan Kejadian Diare. *Jurnal Kesehatan Masyarakat*. 8(2): 176-182.
- Badan Pusat Statistik. 2014. *Potensi Desa Tahun 2014*. Jakarta
- Kemendes RI. 2016. *Profil Kesehatan Indonesia 2015*. Jakarta
- MacQueen, J.B. 1967. "Some Methods for Classification and Analysis of Multivariate Observation". *Proceeding of 5<sup>th</sup> Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297
- Prasetyo, Eko. 2012. *Data Mining, Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta. Penerbit Andi
- Tan, P. et al. 2006. *Introduction to Data Mining*. Boston: Pearson Education