

PENERAPAN ALGORITMA K-MEANS UNTUK MENENTUKAN TINGKAT KESEHATAN BAYI DAN BALITA PADA KABUPATEN DAN KOTA DI JAWA TENGAH

Erga Aprina Sari

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Jl. Imam Bonjol No. 205-207, Semarang, 50131, (024) 3517261
E-mail : aprinaerga@gmail.com

Abstrak

Derajat kesehatan bayi dan balita mencerminkan kesehatan bangsa, sebab bayi dan balita sebagai generasi penerus bangsa memiliki kemampuan yang dapat dikembangkan dalam meneruskan pembangunan bangsa. Berdasarkan alasan tersebut, masalah kesehatan bayi dan balita diprioritaskan dalam perencanaan atau penataan pembangunan bangsa [1]. Salah satu teknik yang dapat digunakan untuk menentukan tingkat kesehatan bayi dan balita adalah dengan menggunakan teknik data mining. Penelitian ini menggunakan algoritma k-means. Dalam penelitian ini menggunakan keseluruhan indikator derajat kesehatan bayi dan balita yaitu angka kematian bayi, angka kematian balita, angka kesakitan dan status gizi. Tujuan dari penelitian ini adalah untuk menghasilkan aplikasi data mining berbasis web dengan algoritma K-Means yang dapat mengklaster kabupaten/kota yang ada di Provinsi Jawa Tengah berdasarkan kemiripan karakteristik daerah yang ditinjau dari nilai indikator kesehatan bayi dan balita sehingga dapat diketahui tingkat derajat kesehatan bayi dan balita untuk setiap kabupaten/kota. Analisa hasil clustering dilakukan pada setiap indikator kesehatan dengan tiga kategori yaitu tinggi, sedang dan rendah. Analisa dilakukan pada tiap indikator agar penanganan dari masalah kesehatan tersebut menjadi lebih terfokus dan tepat sasaran.

Kata kunci: clustering, K-Means, data mining, derajat kesehatan

Abstract

Health status of infants and toddlers reflect the health of the nation, for infants and toddlers as the next generation has the ability that can be developed in the continuing development of the nation. Based on these reasons, the health problems of infants and toddlers are prioritized in planning or arrangement of nation building [1]. One technique that can be used to determine the level of health of infants and toddlers is to use data mining techniques. This study uses k-means algorithm. In this study, using the overall indicator of the health of infants and toddlers are infant mortality, toddlers mortality, morbidity and nutritional status. The purpose of this study is to produce a web-based data mining applications with K-Means algorithm to grouping districts in Central Java province based on similar characteristics of the area in terms of value indicators of the health of infants and toddlers in order to know the level of the health of infants and toddlers to each district / city. Analysis of the results of the clustering is done at each health indicator with three categories: high, medium and low. Analysis performed on each indicator so that the handling of these health problems become more focused and targeted.

Keywords: clustering, K-Means, health status infants and toddlers, data mining

1. Pendahuluan

Terdapat beberapa indikator yang dapat digunakan untuk menentukan derajat kesehatan bayi dan balita di Indonesia yaitu angka kematian bayi, angka kematian balita, angka kesakitan dan status gizi [1]. Tingkat derajat kesehatan bayi dan balita dapat dijadikan

sebagai acuan penting, seperti bagi Pemerintah Daerah (Pemda) agar Pemda lebih terpacu untuk meningkatkan pelayanan kesehatan bayi dan balita sehingga dapat mengurangi angka kematian dan penyakit, serta meningkatkan status gizi bayi dan balita. Selain itu juga sebagai dasar untuk menentukan alokasi dana

bantuan kesehatan dari pemerintah pusat ke daerah.

Dinas Kesehatan Provinsi Jawa Tengah setiap tahunnya melakukan pendataan kesehatan penduduk termasuk bayi dan balita sehingga diperoleh informasi yang berkaitan dengan data kesehatan untuk setiap kabupaten/kota yang ada di Provinsi Jawa Tengah. Namun, Dinas Kesehatan Provinsi Jawa Tengah belum melakukan pengolahan data tersebut, sehingga masing-masing daerah yang ada di Provinsi Jawa Tengah tidak mengetahui tingkat derajat kesehatan bayi dan balita untuk setiap kabupaten/kotanya.

Salah satu teknik yang dapat digunakan untuk menentukan tingkat kesehatan bayi dan balita ini dapat dilakukan dengan teknik *data mining*. *Data mining* dilihat dari sisi teknik pengolahan data menyediakan sejumlah algoritma yang dapat digunakan untuk menggali informasi tersembunyi dari kumpulan data yang multidimensi[2]. Algoritma data mining yang cukup populer digunakan baik dalam dunia bisnis, akademik, ataupun industri adalah algoritma k-Means. Algoritma K-Means merupakan salah satu metode clustering non hirarki yang mempartisi data menjadi beberapa cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [3].

Penelitian ini menggunakan teknik *data mining* dengan algoritma K-Means untuk mengelompokkan atau mengklaster kabupaten-kabupaten yang ada di Provinsi Jawa tengah berdasarkan kemiripan karakteristik daerah yang ditinjau dari nilai indikator kesehatan yaitu angka kematian bayi dan balita, angka kesakitan bayi dan balita, dan status gizi bayi dan balita.

2. Landasan Teori

2.1 Data Mining

Data *mining* dapat diartikan sebagai analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data [4].

2.2 Metode Clustering

Clustering merupakan pengelompokan data, observasi, atau kasus ke dalam kelas dari objek serupa. *Clustering* berbeda dengan *classification*, bahwa tidak ada variabel target clustering. Pengelompokan tidak mencoba untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel sasaran. Sebaliknya, algoritma klasterisasi berusaha untuk mencari bagian dari seluruh dataset menjadi subkelompok yang relatif homogen atau kelompok, di mana kesamaan data dalam cluster dimaksimalkan, dan kesamaan data yang di luar klaster diminimalkan [4].

2.3 Algoritma K-Means

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam beberapa cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [3].

3. Klasterisasi Indikator Kesehatan Bayi dan Balita

3.1 Indikator yang digunakan

Dalam penelitian ini menggunakan indikator mortalitas, morbiditas, dan status gizi:

a. Angka Kematian Bayi

Rumus untuk menghitung angka kematian bayi:

$$IMR = \frac{d_0}{B} \times k \quad (1)$$

Dengan:

d_0 = Jumlah kematian bayi yang berumur kurang dari 1 tahun

B = Jumlah lahir hidup yang dicatat selama 1 tahun

K = Konstanta (1000)

b. Angka kematian balita

Rumus untuk menghitung angka kematian balita :

$$FMR = \frac{b}{B1} \times k \quad (2)$$

Dengan :

b = Jumlah kematian balita yang dicatat selama 1 tahun

B1 = Jumlah penduduk balita pada tahun yang sama

K = Konstanta (1000)

- c. Presentase Balita Gizi Buruk yang dicatat dalam satu tahun.
- d. Presentase Balita Gizi Kurang yang dicatat dalam satu tahun.
- e. Presentase Bayi Berat Lahir Rendah yang dicatat dalam satu tahun.
- f. Presentase Balita Pneumonia yang dicatat dalam satu tahun.
- g. Presentase Balita Diare yang dicatat dalam satu tahun.

3.2 Metode Analisa data

- a. Merubah data ke nilai indikator yang akan digunakan
- b. Normalisasi data

Dalam penelitian ini semua data indikator kesehatan akan dinormalisasikan dalam tingkat 0-1. Berikut ini merupakan rumus untuk normalisasi data [6]:

$$\text{Nilai Baru} = \left(\frac{\text{Nilai asal} - \text{nilai min}}{\text{nilai max} - \text{nilai min}} \right) \quad (6)$$

3.3 Clustering

Langkah awal dalam algoritma K-Means ini tentukan berapa kluster yang akan dibentuk, misal tiga kluster. Kemudian pilih objek yang dijadikan sebagai pusat kluster. hitung jarak setiap data dengan semua titik pusat kluster menggunakan perhitungan *Euclidean Distance*.

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (6)$$

Dengan:

d_{ik} = jarak antara data ke-i dengan titik pusat kluster ke-k

m = jumlah atribut

x_j = data ke-i

c_k = data pusat kluster ke-k

Setelah diperoleh hasil perhitungan jarak setiap data dengan semua titik pusat kluster, data yang memiliki jarak terpendek dengan titik pusat kluster akan menjadi anggota klster tersebut. Selanjutnya, pusat kluster akan diperbaharui berdasarkan nilai rata-rata anggotanya. Hitung jarak setiap data dengan pusat kluster yang baru. Ulangi proses ini sampai nilai pusat kluster tidak berubah.

3.4 Pengujian Kualitas Kluster

Salah satu metode yang sering digunakan untuk pengukuran kualitas kluster adalah Sum of Squared Error [5]. Semakin kecil nilai SSE semakin bagus hasil klustering yang dibuat.

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)^2 \quad (4)$$

4. (4)

BCV adalah rata-rata jarak antar centroid sedangkan WCV adalah Sum of Squared Error. Semakin besar rasio yang dihasilkan maka kualitas kluster semakin bagus[5]. Berikut ini merupakan rumus rasio BCV dan WCV

$$\frac{BCV}{WCV} = \frac{\frac{1}{n_k} \sum_{i=1}^k d(m_i, m_i)}{SSE} \quad (5)$$

4. Analisa dan Pembahasan

Dalam jurnal ini hanya diambil sepuluh sample kabupaten. Sample data yang digunakan dapat dilihat pada tabel 1. Data yang telah hitung menjadi nilai indikator dapat dilihat pada tabel 2. Sebelum dikluster data dinormalisasi terlebih dahulu agar tidak ada parameter yang mendominasi.

Dalam percobaan ini kabupaten/kota yang dipilih sebagai titik pusat adalah Kabupaten Cilacap sebagai titik pusat satu, Kabupaten Purbalingga sebagai titik pusat dua dan Kabupaten kebumen sebagai titik pusat tiga. Iterasi yang terjadi sebanyak tiga kali. Hasil perhitungan titik pusat yang terakhir dapat dilihat pada tabel 4. Kabupaten yang menjadi nggota kluster satu dapat dilihat pada gambar 1, Kabupaten yang menjadi nggota kluster dua dapat dilihat pada gambar 2, Kabupaten yang menjadi nggota kluster tiga dapat dilihat pada gambar 3.

Analisa dilakukan pada hasil perhitungan titik pusat yang terakhir. Kabupaten/kota yang menjadi anggota cluster satu memiliki hasil analisa yang sama dengan analisa pada titik pusat cluster satu. Kabupaten/ kota menjadi anggota cluster dua mempunyai hasil analisa dengan dengan titik pusat cluster dua. Kabupaten/ kota menjadi anggota cluster tiga mempunyai hasil analisa dengan dengan titik pusat cluster tiga. Hasil analisa dapat dilihat pada tabel 5.

Kabupaten/kota yang memiliki hasil analisa indikator kesehatan tinggi merupakan kabupaten/kota yang tingkat kesehatan pada indikator tersebut buruk karena jumlah penderitanya banyak. Penanganan masalah

kesehatan dapat di fokuskan pada kabupaten/kota dengan indikator kesehatan tinggi.

Hasil perhitungan kualitas kluster dapat dilihat pada gambar 4.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari permasalahan di atas dapat disimpulkan bahwa aplikasi *data mining* berbasis web dengan algoritma k-means ini dapat digunakan untuk mengkluster kabupaten/kota yang ada di Provinsi Jawa Tengah berdasarkan kemiripan karakteristik daerah yang ditinjau dari nilai indikator kesehatan bayi dan balita.

Dengan aplikasi berbasis web ini akan mempermudah Dinas Kesehatan Provinsi Jawa Tengah mengetahui tingkat kesehatan bayi dan balita pada setiap kabupaten/kota untuk setiap indikatornya. Analisa tingkat kesehatan bayi dan balita dilakukan pada setiap indikator kesehatan dengan tiga kategori yaitu tinggi, sedang dan rendah.

Analisa dilakukan pada tiap indikator agar penanganan dari masalah kesehatan tersebut menjadi lebih terfokus dan tepat sasaran. Indikator kesehatan yang digunakan adalah angka kematian bayi, angka kematian balita, presentase balita gizi kurang, presentase balita gizi buruk, BBLR, presentase balita pneumonia, dan presentase balita diare.

5.2 Saran

Beberapa saran yang dapat digunakan untuk penelitian selanjutnya adalah sebagai berikut:

1. Indikator untuk morbiditas (kesakitan) yang digunakan dalam penelitian hanya dua penyakit yaitu pneumonia dan diare. Penelitian selanjutnya dapat menambah beberapa penyakit yang sering diderita bayi dan balita seperti demam berdarah dan TBC.
2. Menambahkan visualisasi hasil *clustering* dalam bentuk grafik.
3. Menambahkan media peta sebagai penggambaran hasil distribusi *clustering*.

Daftar Pustaka

- [1] A. A. A. Hidayat, Pengantar Ilmu Kesehatan Anak Untuk Pendidikan Kebidanan, Jakarta, Indonesia: Salemba Medika, 2008.

- [2] P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Boston, USA: Pearson International Edition, 2005.

- [3] Y. Agusta, "K-Means- Penerapan, Permasalahan dan Metode Terkait," Jurnal Sistem dan Informatika Vol.3 : 47-60, Februari 2007.

- [4] D. T. Larose, Discovering Knowledge In Data : An Introduction to Data Mining, Hoboken, New Jersey: John Wiley & Sons, Inc., 2005.

- [5] Arifin, Muljono, S. Sumpeno dan M. Hariadi, "Towards Building Indonesian Viseme: A Clustering-Based Approach".

- [6] T. Mitsa, Temporal Data Mining, Boca Raton: CRC Press, 2010.

Tabel 1 sample beberapa data yang digunakan

tahun	kabupaten kota	jumlah Bayi lahir hidup	jumlah kematian bayi	jumlah kematian balita	jumlah balita	balita gizi kurang(%)	balita gizi buruk(%)	BBLR(%)	balita penderita pneumonia(%)	balita penderita diare(%)
2013	Kab. Cilacap	29652	325	360	136976	3,35	1	4,01	3,52	16
2013	Kab. Banyumas	27744	348	376	136771	1,21	5,26	5,45	14,62	28
2013	Kab. Purbalingga	15214	172	190	68945	2,95	0,81	4,32	5,64	15
2013	Kab. Banjarnegara	16314	271	299	78915	4,76	0,06	6,37	36,01	27
2013	Kab. Kebumen	20112	199	231	11799	0,4	0,03	2,91	86,42	17
2013	Kab. Purworejo	9615	111	132	49780	7,54	0,79	5,26	6,37	13
2013	Kab. Wonosobo	13056	172	184	81935	1,84	0,21	5,42	30,22	43
2013	Kab. Magelang	18993	138	154	93889	1,73	0,04	4,8	16,48	19
2013	Kab. Boyolali	14729	111	120	77271	3,28	0,51	1,54	3,53	3
2013	Kab. Klaten	17734	150	166	92376	3,68	0,63	4,16	20,69	37

Tabel 2 sample perhitungan indikator kesehatan bayi dan balita

tahun	kabupaten kota	balita gizi kurang(%)	balita gizi buruk (%)	BBLR (%)	balita penderita pneumonia (%)	balita penderita diare (%)	angka kematian bayi	angka kematian balita
2013	Kab. Cilacap	3,35	1	4,01	3,52	16	10,96	2,63
2013	Kab. Banyumas	1,21	5,26	5,45	14,62	28	12,54	2,75
2013	Kab. Purbalingga	2,95	0,81	4,32	5,64	15	11,31	2,76
2013	Kab. Banjarnegara	4,76	0,06	6,37	36,01	27	16,61	3,79
2013	Kab. Kebumen	0,4	0,03	2,91	86,42	17	9,89	19,58
2013	Kab. Purworejo	7,54	0,79	5,26	6,37	13	11,54	2,65
2013	Kab. Wonosobo	1,84	0,21	5,42	30,22	43	13,17	2,25
2013	Kab. Magelang	1,73	0,04	4,8	16,48	19	7,27	1,64
2013	Kab. Boyolali	3,28	0,51	1,54	3,53	3	7,54	1,55
2013	Kab. Klaten	3,68	0,63	4,16	20,69	37	8,46	1,8

Tabel 3 sample normalisasi data

tahun	kabupaten kota	balita gizi kurang (%)	balita gizi buruk (%)	BBLR (%)	balita penderita pneumonia (%)	balita penderita diare (%)	angka kematian bayi	angka kematian balita
2013	Kab. Cilacap	0,41	0,19	0,39	0	0,33	0,4	0,06
2013	Kab. Banyumas	0,11	1	0,61	0,13	0,63	0,56	0,07
2013	Kab. Purbalingga	0,36	0,15	0,44	0,03	0,3	0,43	0,07
2013	Kab. Banjarnegara	0,61	0,01	0,76	0,39	0,6	1	0,12
2013	Kab. Kebumen	0	0	0,22	1	0,35	0,28	1
2013	Kab. Purworejo	1	0,15	0,58	0,03	0,25	0,46	0,06
2013	Kab. Wonosobo	0,2	0,03	0,61	0,32	1	0,63	0,04
2013	Kab. Magelang	0,19	0	0,51	0,16	0,4	0	0
2013	Kab. Boyolali	0,4	0,09	0	0	0	0,03	0
2013	Kab. Klaten	0,46	0,11	0,41	0,21	0,85	0,13	0,01

Tahun	Kabupaten/Kota	C1	C2	C3	Cluster
2013	Kab. Cilacap	0.13	0.67	1.46	1
2013	Kab. Banyumas	0.92	1.08	1.72	1
2013	Kab. Purbalingga	0.18	0.65	1.43	1
2013	Kab. Purworejo	0.64	0.88	1.74	1
2013	Kab. Magelang	0.49	0.78	1.38	1
2013	Kab. Boyolali	0.63	1.2	1.55	1

Gambar 1 hasil clustering yang menjadi anggota cluster satu

Tahun	Kabupaten/Kota	C1	C2	C3	Cluster
2013	Kab. Banjarnegara	0.94	0.54	1.55	2
2013	Kab. Wonosobo	0.87	0.29	1.46	2
2013	Kab. Klaten	0.6	0.52	1.46	2

Gambar 2 hasil clustering yang menjadi anggota cluster dua

Tahun	Kabupaten/Kota	C1	C2	C3	Cluster
2013	Kab. Kebumen	1.44	1.41	0	3

Gambar 3 hasil clustering yang menjadi anggota cluster tiga

Tabel 4 hasil perhitungan titik pusat terakhir

tahun	kabupaten kota	balita gizi kurang	balita gizi buruk	BBLR	balita penderita pneumonia	balita penderita diare	angka kematian bayi	angka kematian balita	cluster
2013	Kab. Cilacap	0,41	0,26	0,42	0,06	0,32	0,31	0,04	1
2013	Kab. Purbalingga	0,42	0	0,59	0,31	0,82	0,59	0,06	2
2013	Kab. Kebumen	0	0	0,22	1	0,35	0,28	1	3

Tabel 5 analisa hasil clustering

balita gizi kurang	balita gizi buruk	BBLR	balita penderita pneumonia	balita penderita diare	angka kematian bayi	angka kematian balita	cluster
sedang	tinggi	sedang	rendah	rendah	sedang	rendah	1
tinggi	rendah	tinggi	sedang	tinggi	tinggi	sedang	2
rendah	rendah	rendah	tinggi	sedang	rendah	tinggi	3

Tahun	SSE	BCV	BCV/SSE
2013	2.59	3.55	1.37

Gambar 4 hasil perhitungan BCV/SSE