

PENERAPAN ALGORITMA K-MEANS PADA KUALITAS GIZI BAYI DI INDONESIA

Diajeng Tyas Purwa Hapsari

Teknik Informatika STMIK AMIKOM Yogyakarta
Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55281
Email : diajeng8235@students.amikom.ac.id

Abstrak

Kualitas gizi pada balita mempengaruhi kemajuan bangsa maka dari itu Pemerintah harus mengevaluasi sudah sejauh mana kualitas gizi. Salah satu cara untuk menentukan sejauh mana kualitas gizi bayi adalah dengan data mining algoritma k-means. Dari data yang diolah diketahui sejauh mana kualitas gizi di Indonesia serta untuk keperluan evaluasi pemerintah. Dari data yang diperoleh dan clustering didapatkan hasil pemerintah belum perlu meninjau ulang secara mendalam untuk hal perbaikan gizi.

Kata kunci: data mining, clustering, k-means, gizi balita

1. Pendahuluan

Kesehatan balita sangat berpengaruh pada baik atau buruknya kualitas masyarakat pada suatu daerah maka dari itu perlu adanya pengalokasikan dana yang cukup untuk perbaikan gizi pada balita demi terwujudnya masyarakat yang berkualitas baik. Peningkatan kualitas kesehatan balita juga diimbangi dengan pemberian asi eksklusif, imunisasi rutin serta diadakannya posyandu di tiap-tiap daerah.

Dalam kasus ini dapat ditangani dengan data mining dan Algoritma K-Means. Data mining, sering disebut juga sebagai Knowledge Discovery in Database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data-data yang berukuran besar [2]. Salah satu metode clustering adalah K-Means. K-Means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [3].

Dalam penelitian ini menggunakan data mining dan k means untuk clustering untuk melakukan pengelompokan data berdasarkan tingkat gizi di daerah tersebut.

2. Landasan Teori

2.1. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan mengidentifikasi informasi yang bermanfaat dan

pengetahuan yang terakrit dari berbagai database besar[4].

2.2 Metode Clustering

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan(kemiripan).Clustering beda dengan group, kalau group berarti kelompok yang sama,kondisinya kalau tidak ya pasti bukan kelompoknya.Tetapi kalau cluster tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sample yang ada, salah satunya dengan menggunakan rumus jarak eclidean.Aplikasinya cluster ini sangat banyak, karena hamper dalam mengidentifikasi permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja.

- Identifikasi obyek (Recognition) :
Dalam bidang image Processing , Computer Vision atau robot vision
- Decision Support System dan data mining
Segmentasi pasar, pemetaan wilayah, Manajemen marketing dll.
- Similarity Measures (ukuran kedekatan)
- Distances dan Similarity Coeficients untuk beberapa sepasang dari item
Eclidean Distance:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

Atau :

$$d(x, y) = \left[\sum_{i=1}^p |x_i - y_i|^2 \right]^{1/2}$$

[4].

2.3 Algoritma K-Means

K-Means merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering suatu data. Ada banyak pendekatan untuk membuat cluster, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam group yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah clustering [8]. Metode K-Means adalah metode yang termasuk dalam algoritma clustering berbasis jarak yang membagi data ke dalam

sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik [8]. Pengelompokan data dengan metode K-Means dilakukan dengan algoritma [9]:

1. Tentukan jumlah kelompok
2. Alokasikan data ke dalam kelompok secara acak
3. Hitung pusat kelompok (*centroid*/rata-rata) dari data yang ada di masing-masing kelompok. Lokasi centroid setiap kelompok diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke-i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung centroid fitur ke-i digunakan persamaan 1.

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

persamaan 1 dilakukan sebanyak p dimensi dari i=1 sampai dengan i=p.

4. Alokasikan masing-masing data ke centroid/rata-rata terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, diantaranya adalah Euclidean [9]. Pengukuran jarak pada ruang jarak (*distance space*) Euclidean dapat dicari menggunakan persamaan 2. Dalam metode K-Means didasarkan pada perbandingan jarak antara data dengan centroid setiap kelompok yang ada

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

Pengalokasian kembali ke dalam masing-masing kelompok yang ada[5].

3. Metodologi Penelitian

3.1 Sumber Data

Sumber data dalam penelitian diambil dari dataset pada data.go.id/dataset merupakan data gizi balita yang ada pada setiap provinsi di Indonesia dengan jumlah 133 data dengan field kode provinsi, nama provinsi, status gizi, dan persen gizi.

Berikut adalah data status gizi pada provinsi pada tabel 1.

| kode_provinsi | nama_provinsi | status_gizi | persen_status_gizi |
|---------------|--------------------------------|-------------|--------------------|
| 11 | Prov. Nanggroe Aceh Darussalam | Gizi Buruk | 7.1 |
| 11 | Prov. Nanggroe Aceh Darussalam | Gizi Kurang | 16.6 |
| 11 | Prov. Nanggroe Aceh Darussalam | Gizi Baik | 72.1 |
| 11 | Prov. Nanggroe Aceh Darussalam | Gizi Lebih | 4.2 |
| 12 | Prov. Sumatera Utara | Gizi Buruk | 7.8 |
| 12 | Prov. Sumatera Utara | Gizi Kurang | 13.5 |
| 12 | Prov. Sumatera Utara | Gizi Baik | 71.1 |
| 12 | Prov. Sumatera Utara | Gizi Lebih | 7.5 |
| 13 | Prov. Sumatera Barat | Gizi Buruk | 2.8 |
| 13 | Prov. Sumatera Barat | Gizi Kurang | 14.4 |
| 13 | Prov. Sumatera Barat | Gizi Baik | 81.3 |
| 13 | Prov. Sumatera Barat | Gizi Lebih | 1.6 |
| 14 | Prov. Riau | Gizi Buruk | 4.8 |
| 14 | Prov. Riau | Gizi Kurang | 11.4 |
| 14 | Prov. Riau | Gizi Baik | 75.2 |
| 14 | Prov. Riau | Gizi Lebih | 8.6 |
| 15 | Prov. Jambi | Gizi Buruk | 5.4 |

3.2 Transformasi Data

Pada tahap ini dilakukan proses perubahan data yang bertujuan dapat diolah menggunakan algoritma K-Means Clustering.

Variabel yang dipilih pada data gizi balita yaitu data kode provinsi, status gizi, persen status gizi.

Tabel 2 data telah di preprocessing

| nama_provinsi | status_gizi | persen_status_gizi |
|--------------------------------|-------------|--------------------|
| Prov. Nanggroe Aceh Darussalam | 1 | 1 |
| Prov. Nanggroe Aceh Darussalam | 1 | 1 |
| Prov. Nanggroe Aceh Darussalam | 1 | 1 |
| Prov. Nanggroe Aceh Darussalam | 1 | 2 |
| Prov. Sumatera Utara | 1 | 2 |
| Prov. Sumatera Utara | 2 | 3 |
| Prov. Sumatera Utara | 2 | 3 |
| Prov. Sumatera Utara | 2 | 3 |
| Prov. Sumatera Barat | 2 | 3 |
| Prov. Sumatera Barat | 2 | 3 |
| Prov. Sumatera Barat | 3 | 4 |
| Prov. Sumatera Barat | 3 | 4 |
| Prov. Riau | 3 | 4 |
| Prov. Riau | 3 | 4 |
| Prov. Riau | 4 | 1 |
| Prov. Riau | 4 | 1 |
| Prov. Jambi | 4 | 1 |
| Prov. Jambi | 4 | 2 |
| Prov. Jambi | 4 | 2 |

Variabel status gizi nilai angka 1 diberikan kepada gizi buruk, nilai 2 untuk status gizi kurang, nilai 3 untuk status gizi baik dan status 4 untuk status gizi lebih. Pada variabel status gizi .

Variabel persen gizi

3.3 Pengolahan Data Setelah Proses Transformasi

Langkah selanjutnya adalah proses pengolahan data menggunakan algoritma K-Means Clustering.

1. Tentukan nilai k dari cluster yang ingin dibuat. Cluster yang ingin dibuat adalah 2 cluster.
2. Tentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal ditentukan dengan cara random (acak) yang didapat dari setiap cluster.

Tabel 3 menentukan pusat awal cluster

| Titik Pusat Cluster | nama provinsi | status_gizi | persen_status_gizi |
|---------------------|--------------------------------|-------------|--------------------|
| Cluster1 | Prov. Nanggroe Aceh Darussalam | Buruk | 1 |
| Cluster 2 | Prov Jambi | Gizi Lebih | 2 |

3. Dalam penelitian ini juga menggunakan hard k-means yaitu penemuan titik yang paling dekat dengan data dengan formulasi sebagai berikut:

$$D(1,1)=\sqrt{(1-1)^2+(1-1)^2+(1-1)^2}=0$$

Dari hasil perhitungan di atas maka hasilnya provinsi pertama dengan pusat cluster pertama adalah 0.

4. Jarak data provinsi pertama dengan pusat cluster kedua yaitu

$$D(1,2)=\sqrt{(5-2)^2+(4-2)^2+(2-2)^2}=13$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak antara provinsi pertama dengan pusat cluster ketiga adalah 13.

Berdasarkan hasil kedua perhitungan tersebut bahwa data provinsi pertama yang paling dekat dengan pusat cluster ke 13.

5. Pengelompokan hasil cluster pada iterasi pertama dengan formulasi sebagai berikut

$$C = \frac{\sum m}{n}$$

3. Analisa dan Pembahasan

Pada tahap ini mencoba mengaplikasikan data K-Means Clustering

Initial Cluster Centers

| | Cluster | |
|--------------------|---------|----|
| | 1 | 2 |
| persen_status_gizi | 2 | 81 |

Initial cluster center menunjukkan hasil pengelompokan sementara.

Final Cluster Centers

| | Cluster | |
|--------------------|---------|----|
| | 1 | 2 |
| persen_status_gizi | 8 | 75 |

Dari kedua hasil di atas menunjukkan cluster di atas rata-rata total.

Number of Cases in each Cluster

| | |
|-----------|--------|
| Cluster 1 | 15,000 |
| Cluster 2 | 5,000 |
| Valid | 20,000 |
| Missing | ,000 |

Kesimpulan Cluster

Pada hasil cluster 1 terdapat 15 responden dan cluster 2 terdapat 5 responden

Pada hasil clustering k-means tersebut tingkat gizi pada lima provinsi berada dalam kategori baik. Ditunjukkan dengan initial cluster yang menunjukkan kelompok-kelompok cluster. Final Cluster yang menunjukkan hasil cluster berada pada rata-rata total.

Dan Number of Case in each Cluster menunjukkan terdapat 15 responden pada cluster 1 serta 5 responden pada cluster dua.

Kesimpulan

1. Dari hasil pengamatan K means dapat digunakan untuk penentuan centroid 1 centroid 2 yang dapat disimpulkan bahwa seluruh provinsi memiliki gizi yang baik.
2. Hasil aplikasi yang dirancang dengan metode K-Means Clustering adalah kelompok tingkatan gizi balita di berbagai provinsi di Indonesia.
3. Keluaran dari aplikasi menunjukkan tingkat gizi balita di berbagai provinsi di Indonesia.
4. Dari data yang telah diolah menunjukkan seluruh balita di seluruh provinsi mendapatkan gizi baik
5. Dapat disimpulkan bahwa belum perlu diadakan evaluasi secara lebih mendalam karena tingkat gizi sudah menunjukkan rata-rata yang baik. Sehingga pemerintah belum perlu meninjau ulang perbaikan gizi.

Daftar Pustaka

- [1] (2013) Tips Anak Bayi dan Balita. [Online]. Available: <http://www.duniaanak.org/makanan-anak/pentingnya-gizi-anak-balita.html>[Diakses:16Desember2016].
- [2] Santosa. DATA MINING : Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta. Graha Ilmu. 2007
- [3] Arai, K., Barakbah, A. R.. 2007. Hierarchical K-Means:an algorithm for centroids initialization for K-Means, the Faculty of Science and Engineering, Saga University, Vol. 36, No.1
- [4] Kusriani, Emha, T. L.2009. *Algoritma Data Mining*, Yogyakarta:Penerbit Andi,2009
- [5] Prasetyo, E., 2012, *Data Mining : Konsep dan Aplikasi Menggunakan MATLAB*, Andi, Yogyakarta.

Biodata Penulis

Diajeng Tyas Purwa Hapsari, adalah mahasiswi jurusan Teknik Informatika angkatan 2014 jenjang Strata-1 di STMIK AMIKOM Yogyakarta.

