

Penerapan Metode *K-Means* Untuk *Clustering* Puskesmas Berdasarkan Jumlah Stunting Di Kabupaten Karawang

1st Sigit Gunawan

Universitas Buana Perjuangan
Karawang

Karawang, Indonesia

if18.sigitgunawan@mh.s.ubpkarawang.ac.id

2nd Hanny Hikmayanti

Universitas Buana Perjuangan
Karawang

Karawang, Indonesia

hanny.hikmayanti@ubpkarawang.ac.id

3rd Tohirin Al Mudzakir

Universitas Buana Perjuangan
Karawang

Karawang, Indonesia

tohirin@ubpkarawang.ac.id

Abstract— Kondisi bayi yang mengacu pada kegagalan tahap pertumbuhan dan perkembangan tidak normal dibandingkan dengan anak-anak lain seusianya dikategorikan sebagai stunting. Sejak tahun 2018 Kabupaten Karawang masuk ke dalam 100 Kabupaten/Kota prioritas penanggulangan stunting. Penelitian ini bertujuan untuk memetakan Puskesmas yang terdapat bayi stunting ke dalam 3 kelompok rendah, sedang dan tinggi menggunakan metode *K-means*. Metode *Davies-Bouldin Index* digunakan untuk mengevaluasi hasil *clustering* yang dilakukan *K-means* untuk mendapatkan nilai validitas dari $K=3$. Adapun hasil dari penelitian terhadap data penyebaran stunting di kabupaten karawang *cluster 0* berjumlah 18 Puskesmas, *cluster 2* berjumlah 25 Puskesmas, dan *cluster 3* berjumlah 7 Puskesmas. Hasil *Davies Bouldin Index* pada *python* sebesar 0.9986083164230459 dan manual sebesar 0.97809247, membuktikan bahwa algoritma *K-Means* cukup baik dalam melakukan *Clustering*.

Kata kunci — *Clustering, Davies-Bouldin Index, Stunting*.

I. PENDAHULUAN

Stunting ialah keadaan dimana balita memiliki panjang maupun besar tubuh yang kurang apabila dibandingkan dengan anak seusianya[1]. Keadaan ini diukur dengan panjang tubuh yang lebih dari -2 standar deviasi median dari *World Health Organization*, stunting dianggap sebagai masalah gizi kronik yang disebabkan oleh banyak aspek. Faktor yang dapat menghasilkan status gizi baik adalah pemberian asupan yang seimbang mulai dari pemberian ASI, sanitasi, air bersih, vitamin dan lain-lain. Apabila hal semacam itu tidak diperhatikan maka mengakibatkan ketidakseimbangan gizi yang menyebabkan terjadinya stunting pada anak yang dapat mempengaruhi pola tumbuh kembang anak [2].

Bersumber dari riset kesehatan dasar(RISKESDAS) tahun 2018 terdapat 30,8% balita menderita stunting, jumlah persentasenya sebesar 11,5% anak sangat pendek dan 19,3% anak pendek [3]. Menurut hasil Dinas Pengendalian Penduduk Dan Keluarga Berencana (DPPKB) tahun 2018 Jawa Barat menunjukkan angka 31,06% tersebar di berbagai Kabupaten termasuk Karawang. Sejak tahun 2018 Kabupaten Karawang masuk ke dalam 100 Kabupaten/Kota prioritas penanggulangan stunting sejak tahun 2018. Menurut *World Health Organization* (WHO) permasalahan kesehatan di kalangan masyarakat dikatakan berat apabila kerutinan peristiwa stunting sebesar 30-39 persen dan sangat darurat jika angkanya ≥ 40 persen. Apabila stunting tidak segera ditangani maka akan banyak terlahirnya bayi yang mengalami pertumbuhan inferior ini akan berdampak buruk bagi calon bibit penerus bangsa di akan mendatang. Oleh karena itu salah satu upaya yang bisa dilakukan untuk menekan angka stunting agar *trend* permasalahan gizi termasuk stunting dimasa depan tidak mengalami kenaikan yaitu melakukan pemetaan daerah rawan stunting melalui puskesmas dengan cara mengekstraksi data jumlah balita stunting 5 tahun terakhir berdasarkan jumlah penderita di Puskesmas Karawang dengan teknik *Data Mining*.

Berdasarkan permasalahan tersebut, maka penelitian ini akan mengelompokkan tingkat kejadian stunting pada 5 tahun terakhir di Kabupaten Karawang berdasarkan jumlah penanganan setiap puskesmas menggunakan teknik *Data Mining K-Means clustering* dengan tujuan untuk membagi Puskesmas yang terdapat di kabupaten Karawang kedalam 3 *cluster* yaitu stunting tinggi, sedang dan rendah.

II. METODE PENELITIAN

A. Stunting

Stunting ialah keadaan dimana balita memiliki panjang maupun besar tubuh yang kurang apabila dibandingkan dengan anak seusianya[1]. Keadaan ini diukur dengan panjang tubuh yang lebih dari -2 standar deviasi median dari *World Health Organization*, stunting bisa disebabkan oleh banyak faktor seperti kondisi ekonomi, gizi ibu saat hamil, kesakitan pada anak, serta kekurangan nutrisi pada anak[1]. Dampak yang disebabkan dari masalah gizi stunting ada 2 yaitu jangka panjang dan jangka pendek masing-masing mempunyai resiko yang merugikan bagi penderitanya. Pengawasan khusus perlu didapatkan bagi balita dengan usia 2 sampai 3 tahun, karena usia tersebut cenderung terlambat dalam masa pertumbuhan dibandingkan dengan usia 0-2 tahun [4].

B. Data Mining

Data mining adalah proses mengekstraksi sekumpulan data-data dengan menghasilkan sebuah informasi menarik, yang dapat digunakan secara spesifik guna membantu dalam pengambilan kebijakan. *Data Mining* terdapat banyak metode, yaitu estimasi, klasifikasi, prediksi, *clustering* dan asosiasi. [5].

C. Algoritma K-Means

Algoritma *K-Means* ialah metode *Data Mining* guna mengelompokkan *dataset* ke dalam beberapa partisi, di mana data dalam satu partisi memiliki ciri yang sama satu sama lain dengan data pada partisi lainnya. *K-means* merupakan algoritma yang efisien untuk menemukan kelompok karena caranya memperkecil jarak data ke *cluster*, tumpukan data hasil *cluster* tergantung pada titik pusat *cluster* awal yang ditentukan (*centroid*) [6]. Berikut adalah tahapan-tahapan dalam algoritma *K-Means*:

1. Tentukan jumlah k .
2. Tentukan nilai pusat (*centroid*) dari masing-masing *cluster* secara acak atau random.

$$V = \frac{\sum_{i=1}^n x_i}{n} \quad i = 1, 2, 3, \dots, n \quad (1)$$

Keterangan :

V = Titik pusat *cluster*

x_i = objek data ke- i

n = Jumlah objek data

3. Setiap data di hitung jarak terdekat terhadap *centroid* menggunakan rumus *euclidean distance*.

$$D_{(i,j)} = \sqrt{(x_{1i} + j_{1j})^2 + (x_{2i} + j_{2j})^2 + \dots + (x_{ki} + j_{kj})^2} \quad (2)$$

Keterangan:

$D(i,j)$ = Jarak objek i ke pusat *cluster* j

X_{ki} = Objek ke i pada atribut objek ke k

X_{kj} = Titik pusat ke j pada atribut ke k

4. Kelompokkan data pada *cluster* dan perbaharui lokasi masing-masing pusat *cluster* untuk mendapatkan nilai *centroid* yang baru.
5. Ulangi langkah 3 dan 4 sampai tidak terdapat perpindahan anggota *cluster* ke *cluster* lain, maka iterasi tersebut berhenti.

D. Evaluasi

Evaluasi dilakukan untuk mengetahui nilai validitas atau nilai kualitas terhadap pengelompokan yang telah di hasilkan. Metode *Davies Bouldin Index* adalah metode untuk menghitung validitas dari suatu *cluster* dalam mengelompokkan data berdasarkan skor keterpaduan (kohesi) dan keterpisahan (separasi) [7]. Tahapan dari perhitungan *Davies- Bouldin Index* adalah sebagai berikut:

1. *Sum of Square Within-cluster* (SSW)

Bertujuan untuk mengetahui kohesi dalam sebuah *cluster* ke- i , kohesi ialah jumlah dari jarak objek terhadap titik pusat *cluster* dari sebuah *cluster* yang diikuti. Rumusnya sebagai berikut[7].

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Keterangan:

m_i = Jumlah objek data dalam *cluster* ke- i

x = data dalam *cluster*

$D(x,c)$ = Jarak data *centroid*

X_i = data pada *cluster* tersebut

C_j = Centroid pada *cluster* ke- i

2. *Sum of Square Between-cluster* (SSB)

Mencari *Sum of Square Between-cluster* (SSB) yang bertujuan untuk mendapatkan bobot pemisahan antar *cluster*.

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

Keterangan:

C_i = data pada *cluster* satu

C_j = Centroid pada *cluster* lainnya

3. Ratio (Rasio)

Mencari perbandingan nilai antara *cluster* ke- i dan *cluster* ke- j atau menghitung rasio setiap *cluster*.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

Keterangan:

R_{ij} = rasio antar *cluster*

SSW_i = cluster satu

SSW_j = cluster lainnya

$SSB_{i,j}$ = cluster satu dan lainnya

4. Davies-Bouldin Index

Rumus mencari nilai DBI:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (6)$$

Keterangan:

K = *cluster* yang ada

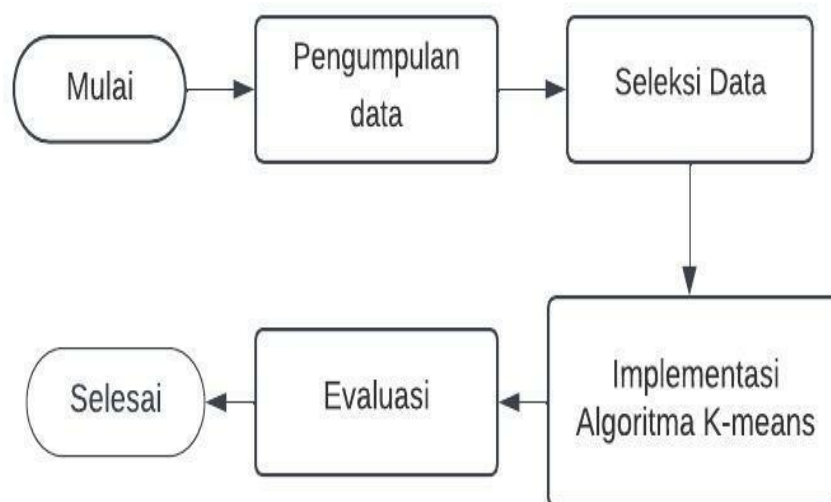
R_{ij} = rasio antar *cluster* i dan j

$\max$ = dicari nilai rasio antar *cluster* yang terbesar

Jumlah *cluster* ialah k , apabila angka DBI yang di dapat rendah (non-negatif ≥ 0), maka semakin bagus *cluster* yang di dapatkan[7].

E. Prosedur Penelitian

Prosedur penelitian dimulai dengan pengumpulan data dengan rincian sebagai berikut:



Gambar 1. Prosedur Penelitian

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data bersumber dari Dinas Kesehatan Kabupaten Karawang yang merupakan data rekap pertahun mengenai anak yang menderita stunting di setiap daerah karawang, data berbentuk *file excel* yang dapat diolah dan diseleksi data tersebut dengan mudah. *Dataset* merupakan tabel data prevalensi bayi yang menderita penyakit stunting dengan kriteria data meliputi 30 nama kecamatan, 50 nama puskesmas, sasaran, pravalensi stunting setiap puskesmas, serta N menggambarkan jumlah pengidap stunting setiap puskesmas di Kabupaten Karawang dari tahun 2018 sampai tahun 2021 yang ditunjukkan tabel 1 sebagai berikut.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	REKAPITULASI HASIL BULAN PENIMBANGAN BALITA															
2	INDIKATOR PB/U ATAU TB/U															
3	KABUPATEN KARAWANG															
4																
5																
6																
7	No	Kecamatan	Puskesmas	INDIKATOR TB/PB ATAU TB/U												
2018				2019			2020			2021						
STUNTING				STUNTING			STUNTING			STUNTING						
9				Sasaran	N	%	Sasaran	N	%	Sasaran	N	%	Sasaran	N	%	
10	1	PANGKALAN	1	PANGKALAN	2777	350	12.6	3002	51	1.7	3196	168	5.3	3117	296	9.5
11	2	TEGALWARU	2	LOJI	2787	108	3.9	2946	48	1.6	2968	39	1.3	2895	15	0.5
12	3	CIAMPEL	3	CIAMPEL	3351	736	22.0	2977	384	12.9	3592	258	7.2	3519	299	8.5
13	4	TELUKJAMBE	4	TELUK JAMBE	5646	240	4.3	4379	17	0.4	4862	209	4.3	5630	112	2.0
14	5	TIMUR	5	WADAS	3113	169	5.4	3336	25	0.7	3393	4	0.1	3645	6	0.2
15	6	WANAKERTA	6	WANAKERTA	3838	369	9.6	4517	63	1.4	4832	62	1.3	4798	51	1.1
16	7	KLARI	7	KLARI	9612	3	0.0	6041	146	2.4	7166	101	1.4	7374	128	1.7
17	8	KLARI	8	CURUG	2478	344	13.9	2538	81	3.2	2569	148	5.8	2285	136	6.0
18	9	KLARI	9	ANGGADITA	1922	64	3.3	1612	169	10.5	1768	226	12.8	1763	199	11.3
19	10	CIKAMPEK	10	CIKAMPEK	7200	12	0.2	7754	26	0.3	8565	179	2.1	7735	76	1.0
20	11	PURWASARI	11	PURWASARI	5399	462	8.6	3828	181	4.7	4369	220	5.0	4493	436	9.7
21	12	TIRTAMULYA	12	TIRTA MULYA	3306	1164	31.6	3155	52	1.6	3212	37	1.2	2926	44	1.5
22	13	JATISARI	13	JATI SARI	2128	256	3.5	2956	248	8.4	3950	42	1.1	3446	62	1.8
23	14	JATISARI	14	PACING	3110	292	21.3	1349	24	1.8	1358	97	7.1	1329	22	1.7
24	15	BANYUSARI	15	CICINDE	3683	267	38.2	975	32	3.3	1736	41	2.4	1525	48	3.1
25	16	BANYUSARI	16	GEMPOL	7385	256	15.6	1587	39	2.5	1731	47	2.7	1563	16	1.0
26	17	KOTABARU	17	KOTA BARU	1373	372	11.3	4518	250	5.5	5126	207	4.0	4680	220	4.7
27	18	KOTABARU	18	JOMIN	699	348	16.4	2220	11	0.5	1856	30	1.6	1833	10	0.5
28	19	CIKAMPEK	19	CIKAMPEK UTARA	1637	157	5.0	2641	236	8.9	2947	118	4.0	2663	97	3.6
29	20	CILAMAYA	20	CILAMAYA	3072	45	1.5	3268	17	0.5	3093	33	1.1	2825	34	1.2
30	21	WETAN	21	SUKATANI	1663	362	21.8	1466	10	0.7	2361	34	1.4	2226	4	0.2
31	22	CILAMAYA	22	PASIRUKEM	2948	921	31.2	2735	38	1.4	2907	18	0.6	2526	21	0.8

Gambar 2. Data Stunting

B. Seleksi Data

1. Seleksi Data

Proses seleksi data merupakan proses penyeleksian terhadap data *excel* yang telah diperoleh sebab tidak semua data dalam *database* diperlukan dalam proses *Data Mining*. Pada awalnya atribut berjumlah 16 atribut setelah dilakukan penyeleksian hanya 5 atribut yang dibutuhkan meliputi atribut Puskesmas dan jumlah balita yang terindikasi stunting tahun 2018, 2019, 2020, 2021. Seperti tabel 2 berikut.

Tabel 1. Data Seleksi

INDIKATOR TB/PB ATAU TB/U					
N0	PUSKESMAS				
		2018	2019	2020	2021
1	Pangkalan	350	51	168	296
2	Loji	108	48	39	15
3	Ciampel	736	384	258	299
4	Teluk Jambe	240	17	209	112
5	Wadas	169	25	4	6
6	Wanakerta	369	63	62	51
7	Klari	3	146	101	128
8	Curug	344	81	148	136
9	Anggadita	64	169	226	199
10	Cikampek	12	26	179	76
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
50	Batujaya	666	46	11	8

2. Pembersihan Data

Proses selanjutnya pembersihan data untuk menghilangkan serta meminimalisir adanya duplikat data, ketidak konsistenan data atau data yang kurang relevan serta meluruskan kesalahan pada data. Pada data stunting 2018-2021 dari Dinas Kesehatan Kabupaten Karawang yang terdiri dari 50 data puskesmas, hasilnya tidak ditemukan *Missing Value* dan duplikat data.

3. Transformation Data

Transformasi data mengubah data asli kedalam bentuk lain yang cocok dengan kebutuhan *Data Mining*. *Dataset* stunting awalnya berbentuk format *Xls* di transformasikan ke dalam format *Csv* untuk menyesuaikan kebutuhan dalam pemrograman data mining pada *python*.

C. Implementasi K-Means

1. Perhitungan Manual

Proses perhitungan *K-Means* dengan pembentukan $K=3$ menggunakan *excel* 2019 yang dilakukan 10 kali itersi menghasilkan *cluster* 0(rendah) berjumlah 18 puskesmas, *cluster* 1(sedang) berjumlah 25 puskesmas dan *cluster* 2(tinggi) berjumlah 7 puskesmas .

2. Program Python

Berdasarkan perhitungan algoritma *K-Means* melalui sistem dengan aplikasi *Jupyter Notebook* dengan menentukan pemodelan *cluster* sebanyak $K=3$ serta *random state* = 16. Hasil *clustering* dapat dilihat pada gambar di bawah ini.

```
# Menampilkan Hasil Kluster
print(kmeanModel.labels_)

[1 0 2 1 0 1 0 1 0 0 1 2 1 1 0 0 1 1 0 0 1 2 0 2 1 1 1 1 0 1 1 0 0 0 1 0 1
 0 1 2 1 1 1 1 0 2 1 1 2 0]
```

```
# Menambahkan Kolom "kluster" Dalam Data Frame
df["kluster"] = kmeanModel.labels_
df
```

	Tahun2018	Tahun2019	Tahun2020	Tahun2021	kluster
0	350	51	168	296	1
1	108	48	39	15	0
2	736	384	258	299	2
3	240	17	209	112	1
4	169	25	4	6	0
5	369	63	62	51	1
...	...	...	...	...	...
45	883	36	15	7	2
46	577	47	62	18	1
47	594	40	7	151	1
48	666	46	11	8	2
49	100	34	3	9	0

Gambar 3. Hasil Cluster

D. Evaluasi

Evaluasi dilakukan untuk mengetahui nilai kualitas dari *cluster* yang telah dibentuk sebanyak 3 *cluster* menggunakan *Davies Bouldin Index*.

1. Perhitungan DBI Manual

Perhitungan manual evaluasi *Davies Bouldin Index* terhadap pengelompokan algoritma *K-Means* menghasilkan nilai DBI sebesar 0.97809247.

2. Perhitungan DBI Python

Berdasarkan evaluasi *Davies Bouldin Index* menggunakan sistem didapatkan hasil dbi seperti gambar 4.

```
from sklearn.metrics import davies_bouldin_score

db_index = davies_bouldin_score(stunting_array, kmeanModel.fit_predict(stunting_array))
print(db_index)
```

0.9986083164230459

Gambar 4. Evaluasi *Davies Bouldin Index* Sistem

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian menghasilkan pengelompokan terhadap dataset dengan pembentukan *cluster* sebanyak $K=3$ yang dilakukan dengan 10 kali iterasi. menunjukkan *cluster* 0(rendah) berjumlah 18 puskesmas, *cluster* 1(sedang) berjumlah 25 puskesmas dan *cluster* 2(tinggi) berjumlah 7 puskesmas. Dari hasil evaluasi terhadap kualitas *cluster* dengan menggunakan *Davies Bouldin Index* pada *python* mendapat hasil sebesar 0.9986083164230459. Sedangkan manual mendapatkan hasil 0.97809247, terjadi perbedaan hasil yang tidak terlalu signifikan hal ini dapat terjadi akibat faktor *Human Error* atau faktor lainnya. Akan tetapi dari hasil evaluasi keduanya membuktikan bahwa algoritma *K-Means* cukup baik untuk melakukan *clustering*.

Pada penelitian ini masih memiliki banyak kekurangan. Untuk itu diharapkan penelitian ini kedepannya dapat dikembangkan lebih lanjut dari aspek dukungan data, penggunaan algoritma dan penggunaan aplikasi lain selain *anaconda*, *excel*.

PENGAKUAN

Naskah ilmiah ini adalah sebagai dari penelitian Tugas Akhir milik Sigit Gunawan, dengan Judul Implementasi Algoritma *K-Means* untuk Pengelompokan Puskesmas Berdasarkan Penanganan Jumlah Stunting Di Kabupaten Karawang yang dibimbing oleh Dr. Hanny Hikmayanti. H, M.Kom dan Tohirin Al Mudzakir, M.Kom.

DAFTAR PUSTAKA

- [1] T. D. Harjanto, A. Vatesia, and R. Faurina, "Analisis Penetapan Skala Prioritas Penanganan Balita Stunting Menggunakan Metode DBScan Clustering," *J. Rekursif*, vol. 9, no. 1, pp. 30–42, 2021.
- [2] C. R. Ardianti, "Implementasi Metode Fuzzy C-Means Untuk Klasifikasi Status Gizi Pada Balita Berdasarkan Indeks Antropometri," *Progr. Stud. Tek. Inform. Fak. Teknol. Inf. dan elektro Univ. Teknol. yogyakarta 2019*, 2019.
- [3] Kemenkes RI, "Buletin Stunting," *Kementeri. Kesehat. RI*, vol. 301, no. 5, pp. 1163–1178, 2018.
- [4] M. Y. Titimeidara and W. Hadikurniawati, "Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita," *J. Ilm. Inform.*, vol. 09, pp. 54–59, 2021.
- [5] R. Kandayu, W. Silaban, M. Maulana, and E. S. Ompusunggu, "Koperasi Dengan Metode Decision Tree," vol. 5, no. 1, pp. 19–23, 2021.
- [6] M. Fauzi, "Penerapan Algoritma K-Means Clustering Untuk Mendeteksi Penyebaran Penyakit TBC (Studi Kasus: Di Kabupaten Deli Serdang)," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 1, no. 2, 2018, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/download/41/38>.
- [7] M. A. Fajar, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means," *J. Tek. Inform.*, pp. 1–5, 2015, [Online]. Available: http://eprints.dinus.ac.id/16498/1/jurnal_15441.pdf (05 Juni 2020).