

Penentuan Kelompok Status Gizi Balita dengan Menggunakan Metode K-Means

Oktania Purwaningrum¹, Yudha Yunanto Putra², Amalia Anjani Ariffiyanti³

^{1,2,3}Universitas Pembangunan Nasional “Veteran” Jawa Timur

¹18082010029@student.upnjatim.ac.id, ²18082010032@student.upnjatim.ac.id,

³amalia_anjani.fik@upnjatim.ac.id

ABSTRAK. Kekurangan gizi pada balita merupakan permasalahan yang masih sering terjadi di Indonesia. Menurut data dari Badan Pusat Statistik (BPS) menunjukkan prevalensi balita kekurangan gizi pada tahun 2018 diatas nilai 10 dan Jawa Timur memiliki nilai 15,20. Permasalahan gizi balita memberikan dampak buruk pada kesehatan balita. Untuk menangani permasalahan kesehatan dan gizi balita hadir adanya posyandu. Posyandu adalah kegiatan yang dilakukan oleh dan untuk masyarakat dibantu oleh petugas kesehatan setempat atau Puskesmas, salah satunya Posyandu Tanjung 1 RW 06 Desa Pepelegi Sidoarjo. Untuk mengelompokan status gizi balita pada Posyandu Tanjung 1 RW 06 Desa Pepelegi Sidoarjo, dilakukan klusterisasi agar penanganan gizi ke balita sesuai sasaran. Proses klusterisasi dilakukan dengan *framework* CRISP-DM dan algoritma *K-Means*. Hasil dari analisis klusterisasi menunjukkan bahwa terdapat 3 klaster dari data balita. Diharapkan kader Posyandu menjadikan hasil tersebut menjadi dasar dalam penentuan kebijakan dalam pemberian tindakan seperti vaksinasi dan imunisasi agar sesuai dan tepat guna.

Kata Kunci: *gizi balita, klusterisasi, CRISP-DM, K-Means*

ABSTRACT. Malnutrition in children under five is a problem that still often occurs in Indonesia. According to data from the Central Statistics Agency (BPS) shows the prevalence of malnourished children under five in 2018 is above 10 and East Java has a value of 15.20. Nutritional problems for toddlers have a bad impact on toddler health. To deal with health and nutrition problems for toddlers, there is a Posyandu. Posyandu is an activity carried out by and for the community assisted by local health workers or Puskesmas, one of which is Posyandu Tanjung 1 RW 06 Desa Pepelegi, Sidoarjo. To classify the nutritional status of toddlers at Posyandu Tanjung 1 RW 06 Desa Pepelegi, Sidoarjo, clustering was carried out so that the handling of nutrition for toddlers was on target. The clustering process is carried out using the CRISP-DM framework and the K-Means algorithm. The results of the clustering analysis show that there are 3 clusters of toddler data. It is hoped that the Posyandu cadres will make these results the basis for determining policies in providing actions such as vaccination and immunization to be appropriate and effective.

Keywords: *toddler nutrition, clustering, CRISP-DM, K-Means*

1. PENDAHULUAN

Gizi adalah zat makanan dibutuhkan untuk perkembangan, pertumbuhan dan kesehatan tubuh seseorang. Gizi seimbang merupakan zat gizi yang mengandung susunan makanan sehari-hari pada porsi dan jenis yang sesuai dengan kebutuhan tubuh yaitu status kesehatan, jenis kelamin, dan umur. Pola makan tidak baik atau tidak bergizi seimbang mempunyai resiko yaitu terjadinya kekurangan gizi seperti berat badan kurang dan anemia. Gizi berlebih (obesitas) juga merupakan resiko yang mungkin terjadi. Lalu penyakit lainnya seperti penyakit jantung koroner, hipertensi, dan diabetes juga mengintai kesehatan tubuh jika gizi tidak diimbangi dengan baik (Handoko, 2020).

Masalah yang sering terjadi pada balita adalah terkait gizi dan tumbuh kembang balita. Menurut Rahim pada Adhimas, gizi buruk yang terjadi pada balita sering disebabkan dari faktor gizi makanan yang balita dapatkan (Nalendra, 2018). Permasalahan gizi balita selalu menjadi sorotan pemerintah agar dapat berbenah pada pemenuhan gizi balita. Data Badan Pusat Statistik (BPS) menunjukkan bahwa prevalensi balita yang kekurangan gizi pada tahun 2018 diatas nilai 10, dan Jawa Timur memiliki nilai 15,20 (Statistik, 2018). Hal tersebut juga menjadi perhatian dari orang tua agar memiliki pengetahuan terkait kandungan gizi dari makanan. Para orang tua harus pandai-pandai dalam memberikan asupan pada balita, agar balita memiliki asupan makanan yang seimbang. Adhimas menyatakan bahwa para orang tua di Indonesia kurang memiliki pengetahuan terkait kandungan gizi pada makanan anaknya. Hal ini disebabkan oleh berbagai faktor seperti: pendidikan, ekonomi, lingkungan, dan kepedulian terhadap sesama masyarakat (Nalendra, 2018).

Permasalahan gizi pada balita memberikan dampak buruk pada kesehatan balita. Menurut Windha dan Intan & Lilik, gizi yang kurang pada balita menghambat pertumbuhan fisik dan mental dari balita selanjutnya akan berpengaruh pada perkembangan balita seperti berjalan, berbicara, belajar, makan, dan lain-lain. Kekurangan gizi juga menyebabkan kecerdasan intelektual (IQ) pada ballita cenderung rendah dan menurunnya

daya tahan tubuh yang menyebabkan berbagai penyakit diare, ISPA (Infeksi Saluran Pernapasan), dan Tuberculosis (Alpiana & Anifah, 2019; Dhuhita, 2015).

Untuk menangani permasalahan kesehatan dan gizi balita hadir adanya posyandu. Menurut Eni & Siti, Posyandu adalah kegiatan yang dilakukan oleh dan untuk masyarakat yang dibantu oleh petugas kesehatan setempat atau Puskesmas (Irfiani & Rani, 2018). Posyandu merupakan salah satu bentuk upaya penyelenggaraan pembangunan kesehatan. Dalam hal ini pemerintah melibatkan masyarakat dan memberikan kemudahan kepada masyarakat dalam memperoleh pelayanan kesehatan dasar yang merupakan salah satu cara untuk mempercepat penurunan Angka Kematian Bayi dan Angka Kematian Ibu. Posyandu sendiri adalah wadah masyarakat untuk memberdayakan diri yang dibentuk berdasarkan musyawarah mufakat pada desa/kelurahan. Posyandu sendiri dikelola oleh pengelola Posyandu, dan dikukuhkan dengan dasar hukum keputusan kepala desa/lurah setempat (Aceh, 2020).

Teknik klastering atau *cluster analysis* dapat digunakan dalam mengelompokkan status gizi balita yang ada pada Posyandu Tanjung 1 RW 06 Desa Pepelegi Sidoarjo (Aprilia et al., 2019; Sitohang & Rikki, 2019). Klasterisasi nilai gizi balita pada Posyandu Tanjung 1 RW 06 Desa Pepelegi Sidoarjo menggunakan metode *K-Means* merupakan hal yang dilakukan pada penelitian ini. Penelitian ini menggunakan data berat badan, tinggi badan, dan umur (bulan) dari balita yang merupakan indikator dari status gizi balita. Hasil analisis kluster diharapkan dapat membantu kader posyandu dalam pembuatan kebijakan yang tepat sasaran terhadap masing-masing kelompok status gizi balita.

2. METODOLOGI

Metodologi penelitian ini merujuk pada *framework* CRISP-DM. Menurut Dina & Dadan, *Cross-Industry Standard Process for Data Mining/* CRISP-DM adalah standar yang digunakan dalam *data mining* untuk memecahkan permasalahan strategis secara umum baik di unit bisnis atau penelitian (Feblian & Daihani, 2017). Menurut Adimas dan Dina & Dadan, CRISP-DM memiliki enam siklus hidup (Feblian & Daihani, 2017; Nalendra, 2018). Enam siklus tersebut digunakan pada penelitian ini yaitu sebagai berikut.



2.1 Pemahaman Bisnis

Fase ini meliputi melihat dan menilai situasi dari kondisi bisnis dan masalah yang ada. Permasalahan gizi balita yang membutuhkan perhatian lebih namun perlu tindakan yang efisien dalam penanganannya dapat dilakukan jika diketahui kelompok-kelompok status gizinya. Pembentukan kelompok ini dapat diselesaikan dengan metode analisis kluster. Tujuannya adalah melakukan klastering nilai gizi pada Posyandu Tanjung 1 RW 06 Desa Pepelegi Sidoarjo. Metode klastering yang dipilih adalah K-Means.

2.2 Pemahaman Data

Setelah tujuan telah ditentukan, maka data mulai dikumpulkan. Penelitian menggunakan data pada Posyandu Tanjung 1 RW 06 Desa Pepelegi Sidoarjo. Atribut data yang digunakan dalam pengelompokkan status gizi adalah dengan menggunakan data berat badan, tinggi badan, dan umur berdasarkan bulan pada Februari 2018. Total data yang digunakan sebanyak 82 data yang berasal dari 82 balita yang berbeda. Pada Tabel 1 ditunjukkan 10 contoh data yang akan digunakan pada penelitian ini.

Tabel 1. Contoh Data Berat Badan, Tinggi Badan, dan Umur Balita

Nomor	Berat Badan (kg)	Tinggi Badan (cm)	Umur (bulan)
1	11.3	84	23
2	7.3	65	14
3	6.9	58	9
4	10.9	92	40
5	8.2	65	8

Nomor	Berat Badan (kg)	Tinggi Badan (cm)	Umur (bulan)
6	16	89	24
7	9.2	63	10
8	8.2	74	13
9	8.6	76	13
10	9.6	76	9

Tabel 2. Eksplorasi Statistik *Dataset*

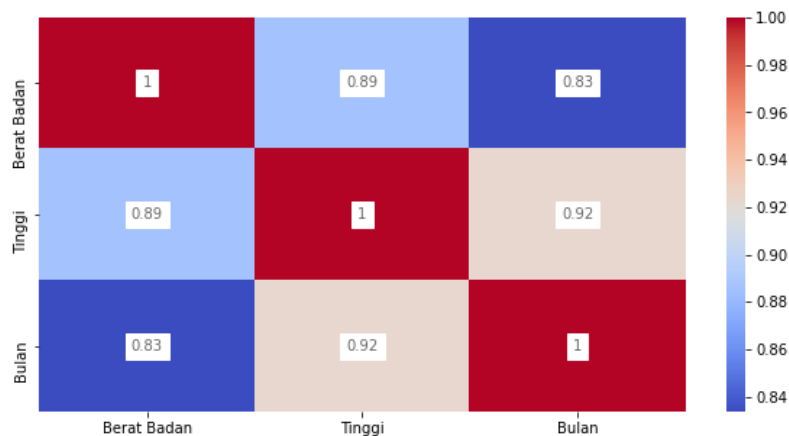
	Berat Badan (kg)	Tinggi Badan (cm)	Umur (bulan)
Count	82.000000	82.000000	82.000000
Mean	11.704878	85.750000	27.646341
Std	3.623514	15.306387	15.500197
Min	6.200000	58.000000	4.000000
25%	8.850000	72.250000	14.250000
50%	11.150000	88.000000	24.000000
75%	14.725000	98.000000	40.000000
Max	22.000000	120.000000	60.000000

Tabel 3. Tipe Data Kolom *Dataset*

Kolom	Tipe Data
Berat Badan	float64
Tinggi	float64
Bulan	int64

Tabel 4. Pengecekan *Null Values Dataset*

Kolom	Null
Berat Badan	0
Tinggi	0
Bulan	0



Gambar 1. Korelasi antar atribut

Dari hasil eksplorasi data tersebut (Tabel 2) diketahui bahwa jumlah kolom dan baris dari *dataset* yaitu terdapat 3 kolom: Berat Badan, Tinggi, dan Bulan. Selain itu jumlah total *instance* berjumlah 82 *instance* (data balita). Pada tabel 2 diketahui pula tipe data dari masing-masing kolom yaitu: kolom Berat Badan bertipe *float*,

Tinggi bertipe *float*, dan Bulan bertipe *int*. Diketahui pula mengenai statistik dasar dari masing-masing kolom yang ada pada dataset yaitu seperti jumlah baris (*count*), rata-rata (*mean*), standar deviasi (*std*), dll. Pada *dataset* yang digunakan pada penelitian ini tidak terdapat data kosong/null pada setiap kolomnya seperti yang terlihat pada tabel 4. Pada Gambar 1 dapat diketahui bahwa ketiga atribut memiliki hubungan korelasi positif yang cukup kuat dengan nilai yang mendekati 1. Dapat dilihat dari korelasi antara berat badan dan tinggi badan yaitu 0.89, korelasi antar berat badan dan umur yaitu 0.83, serta korelasi antara tinggi badan dan umur yaitu 0.92.

2.3 Penyiapan Data

Pada tahap ini, data yang telah dikumpulkan kemudian diolah untuk menangani data kosong dan juga memilih atribut yang digunakan pada tahap pemodelan serta melakukan konversi bentuk atribut agar dapat diolah pada tahap pemodelan. Pada data yang telah dikumpulkan tidak terdapat data yang kosong dan tidak terdapat data yang dibuang.

Pada tahap ini juga ditentukan jumlah kluster yang akan dibuat dalam tahap pemodelan. Penentuan jumlah kluster ini menggunakan dua metode yaitu metode *Elbow* dan metode *Silhouette Coefficient*.

2.4 Pemodelan

Fase ini melakukan pemilihan dan mengaplikasikan model yang sesuai, mengkalibrasikan aturan pada model untuk mendapatkan hasil yang optimal. Pada penelitian ini menggunakan algoritma *K-Means*. Metode *K-Means* adalah teknik penyekatan (*partisi*) yang memisahkan atau membagi objek yang berbeda, dan mengumpulkan objek yang sama ke dalam jumlah *K* daerah yang telah ditentukan sebelumnya dan setiap objek harus masuk dalam kelompok tersebut. Berikut merupakan tahap dari Algoritma *K-Means* (Fahik et al., 2018):

1. Menentukan banyak *K cluster* yang ingin dibentuk.
2. Membangkitkan nilai random untuk pusat awal *cluster* sebanyak *K cluster*.
3. Menghitung jarak setiap data terhadap masing-masing *centroid* menggunakan rumus *euclidean distance*.
4. Mengelompokkan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).
5. Memperbaharui nilai *centroid* berdasarkan dari rata-rata cluster dengan persamaan:

$$C_k = \frac{1}{nk} \sum d_i \quad (1)$$

Dengan:

C_k = *cluster*

Nk = jumlah data dalam cluster

d_i = jumlah tiap objek yang masuk dalam masing-masing *cluster*

6. Ulangi langkah 2 sampai langkah ke 5 hingga anggota tiap cluster tidak ada yang berpindah.

2.5 Evaluasi

Fase ini yaitu melakukan evaluasi model yang telah digunakan untuk melihat kualitas dan efektivitas dari model yang digunakan, mengambil keputusan dari hasil yang dihasilkan, dan meninjau model apa yang memenuhi tujuan. Pada penelitian ini tidak melakukan evaluasi karena tidak ada nilai pembandingan dan pada *clustering* hanya bertujuan untuk mengelompokkan data bukan untuk melabeli data seperti pada metode *classification* (Kaggle, 2020; Sari, n.d.; Sutanto, n.d.).

2.6 Deployment

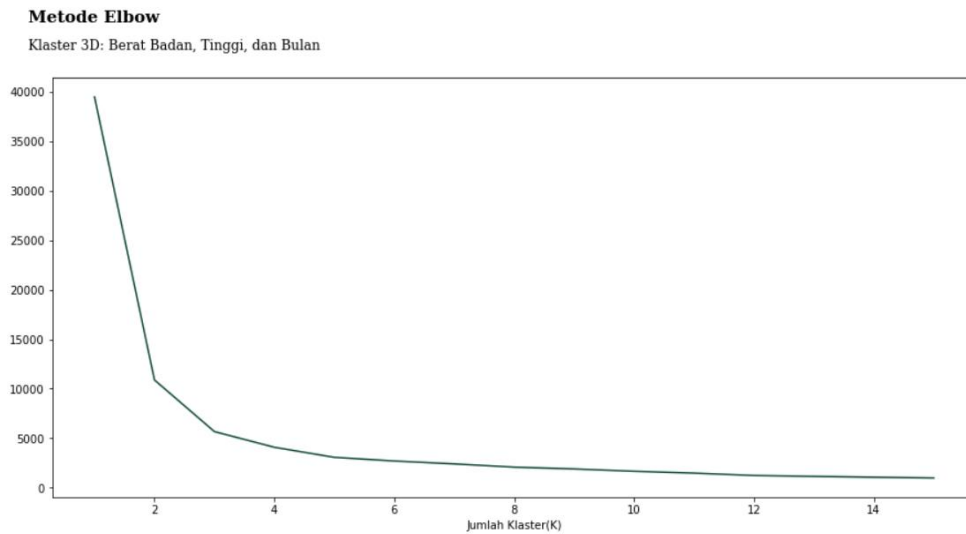
Fase ini melakukan implementasi model yang telah dihasilkan dan dievaluasi dari tahap sebelumnya. Implementasi model yaitu dengan cara menentukan tiap instance yaitu tiap balita termasuk pada kluster tertentu. Setiap balita dalam satu kluster yang sama akan memiliki kemiripan nilai dari tiga atribut yang dianalisis yaitu tinggi badan, berat badan, dan umur.

3. HASIL DAN PEMBAHASAN

3.1 Penentuan Jumlah Kluster

Dalam melakukan klustering, jumlah kluster merupakan hal yang penting. Hal ini dikarenakan banyaknya kluster dapat menentukan hasil akhir dan sikap apa yang dapat dilakukan dalam mengambil suatu

keputusan nantinya. Untuk menentukan jumlah kluster pada penelitian ini digunakanlah metode *Elbow*. Berikut merupakan hasil dari perhitungan metode *Elbow* (Gambar 2).



Gambar 2. Grafik Metode *Elbow*

Berdasarkan hasil metode elbow diatas, dapat disimpulkan bahwa dari data yang dimiliki memiliki 3 kluster. Untuk melakukan validasi dari jumlah kluster yang akan dipilih digunakan metode *Silhouette Coefficient* yaitu mengukur seberapa baik jumlah kluster dengan berdasarkan jarak antar kluster dengan antar entitas dalam kluster. *Silhouette Coefficient* berentang dari -1 hingga +1, dengan nilai mendekati +1 semakin baik kluster tersebut. Berikut merupakan hasil dari perhitungan metode *Silhouette Coefficient* (Gambar 3).

```
Jumlah kluster = 3 nilai rata-rata silhouette = 0.5144196446258764
Jumlah kluster = 4 nilai rata-rata silhouette = 0.4426592053371997
Jumlah kluster = 5 nilai rata-rata silhouette = 0.4145732212502591
Jumlah kluster = 6 nilai rata-rata silhouette = 0.37729625729930744
Jumlah kluster = 7 nilai rata-rata silhouette = 0.35349786214077644
```

Gambar 3. Hasil Perhitungan Metode *Silhouette Coefficient*

Dari hasil metode *Silhouette Coefficient* dapat disimpulkan bahwa jumlah kluster 3 lebih baik digunakan pada penelitian ini karena memiliki nilai mendekati 1 daripada jumlah kluster yang lain, yaitu dengan nilai 0.514.

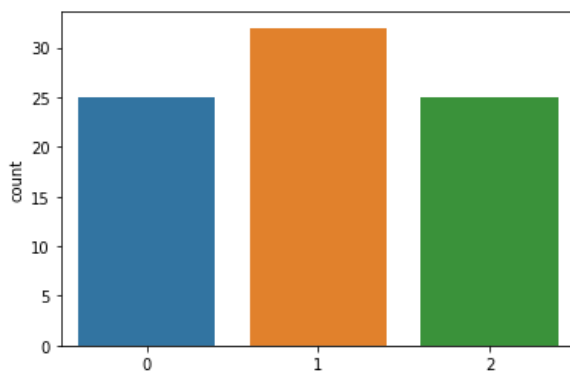
3.2 Pemodelan – Pembentukan Kluster dengan K-Means

Setelah melakukan perhitungan metode *Elbow* dan metode *Silhouette Coefficient* telah diketahui jumlah kluster yang akan muncul, yaitu menggunakan jumlah kluster 3. Maka langkah selanjutnya yang ditempuh yaitu melakukan pengelompokan data sesuai dengan klusternya. Kelompok kluster suatu data ditentukan dari jarak terdekat kelompok tersebut dengan *centroid* atau titik pusat dari suatu kluster data. Titik centroid tiap kluster ditunjukkan pada Tabel 5.

Tabel 5. Titik Centroid tiap Kluster

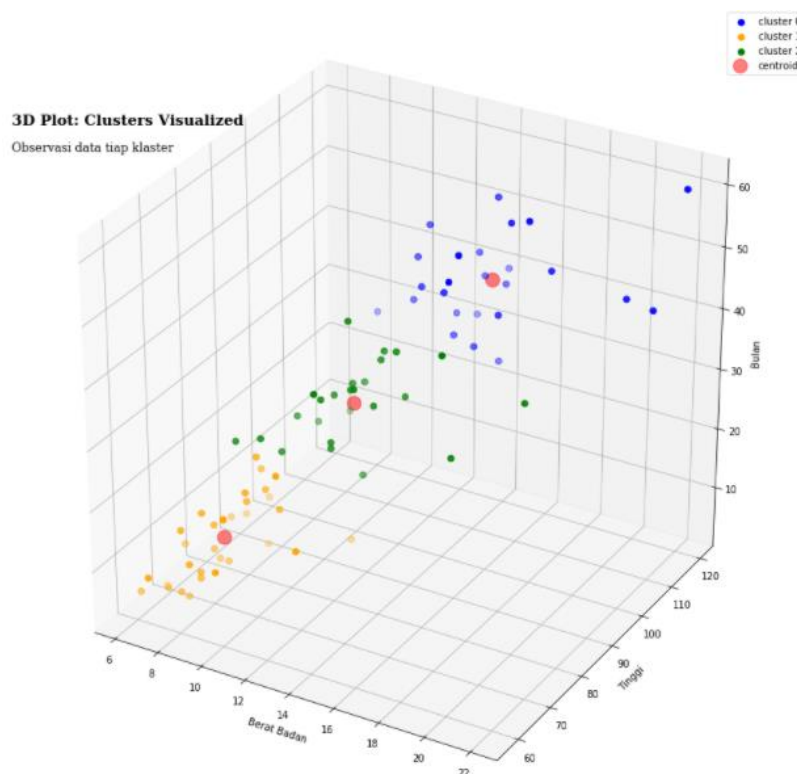
Cluster	Berat Badan (kg)	Tinggi Badan (cm)	Umur (bulan)
0	15.8	103.1	46.7
1	8.6	69.5	12.1
2	11.6	89.1	28.5

Jumlah data dari tiap kluster ditunjukkan pada Gambar 4, dari gambar tersebut dapat disimpulkan jumlah data yang paling banyak ada di kluster 1.



Gambar 4. Jumlah Data tiap Klaster

Sedangkan visualisasi data tiap klaster ditampilkan pada Gambar 5. Dari visualisasi klaster tersebut dan centroid dapat dinyatakan bahwa klaster 0 merupakan kelompok balita dengan rentang usia paling tua dibanding usia balita pada kluster lainnya dengan berat dan tinggi badan yang tinggi. Sedangkan klaster 1 merupakan kelompok balita dengan rentang usia paling muda, berat badan paling ringan, serta tinggi paling pendek jika dibandingkan dengan kluster lainnya. Lalu untuk klaster 2 mempunyai rentang umur, berat badan, dan tinggi yang terletak diantara 2 kluster lainnya yaitu klaster 0 dan klaster 1.



Gambar 5. Visualisasi Data Klaster

3.3 Penetapan Klaster

Setelah dibentuk klusternya, maka tahap berikutnya adalah mencari tahu suatu data termasuk ke dalam klaster mana. Hal ini dilakukan agar tindakan yang dilakukan berdasarkan setiap kluster sesuai dengan anggota dalam klaster tersebut. Contoh penetapan klaster untuk 10 data balita dapat dilihat pada tabel berikut.

Tabel 6. Hasil Klaster pada 10 Data Balita

Nomor	Berat Badan (kg)	Tinggi Badan (cm)	Umur (bulan)	Cluster
1	11.3	84	23	2
2	7.3	65	14	1
3	6.9	58	9	1
4	10.9	92	40	2
5	8.2	65	8	1
6	16	89	24	2
7	9.2	63	10	1
8	8.2	74	13	1
9	8.6	76	13	1
10	9.6	76	9	1

4. KESIMPULAN

Penelitian ini melakukan klasterisasi pada gizi pada balita di Posyandu Tanjung 1 RW 06 Desa Pepelegi Sidoarjo dengan menggunakan algoritma K-Means dengan framework CRISP-DM dengan menghasilkan 3 klaster yang berasal dari perhitungan metode Elbow dan metode Silhouette Coefficient. Dari ketiga klaster yang dihasilkan, klaster 1 memiliki anggota yang jumlahnya paling tinggi dibandingkan klaster 0 dan 2. Klaster 1 memiliki anggota sebanyak 32 balita dengan rentang nilai pada berat badan, tinggi badan, dan umur relatif lebih kecil daripada klaster lain. Klaster 0 memiliki anggota 25 balita dengan rentang nilai berat badan, tinggi badan, dan umur relatif lebih besar daripada klaster lain. Dan klaster 2 memiliki anggota 25 balita dengan rentang nilai berat badan, tinggi badan, dan umur diantara nilai dari klaster 0 dan 1. Dari hasil klaster tersebut, kader posyandu dapat menjadikan hasil temuan sebagai acuan dalam memberikan pelayanan kepada balita. Sehingga pelayanan kepada balita dapat lebih tepat sasaran dan maksimal lagi.

Disarankan pembentukan klaster dihitung kembali dalam rentang periode tertentu, sehingga data yang diobservasi dan klaster yang terbentuk merupakan hasil yang relevan dengan kondisi terbaru.

5. DAFTAR RUJUKAN

- Aceh, D. (2020). *Apa itu Posyandu?* <https://dinkes.acehprov.go.id/news/read/2020/01/09/109/apa-itu-posyandu.html#:~:text=Posyandu merupakan salah satu bentuk,dalam memperoleh pelayanan kesehatan dasar%2F>
- Alpiana, I., & Anifah, L. (2019). Penerapan Metode KnA (Kombinasi K-Means dan Agglomerative Hierarchical Clustering) dengan Pendekatan Single Linkage untuk Menentukan Status Gizi pada Balita. *Indonesian Journal of Engineering and Technology (INAJET)*, 1(2). <https://journal.unesa.ac.id/index.php/inajet>
- Aprilia, A., Rahmawati, W. M., & Hakimah, M. (2019). Penentuan Kategori Status Gizi Balita Menggunakan Penggabungan Metode Klasterisasi Agglomerative Dan K-Means. *Seminar Nasional Sains Dan Teknologi Terapan VII 2019, Institut Teknologi Adhi Tama Surabaya, VII*, 595–600.
- Dhuhita, W. M. P. (2015). Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita. *Jurnal Informatika*, 15(2), 160–174.
- Fahik, B. Y. L., Djahi, B. S., & Rumlaklak, N. D. (2018). Data Mining untuk Klasifikasi Status Gizi Desa di Kabupaten Malaka Menggunakan Metode K-Nearest Neighbor. *J-Icon Jurnal Komputer & Informatika*, 6(1), 1–7.
- Feblian, D., & Daihani, D. U. (2017). Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X. *Jurnal Teknik Industri*, 6(1), 1–12. <https://doi.org/10.25105/jti.v6i1.1526>
- Handoko, S. (2020). *Gizi Seimbang Untuk Gaya Hidup Yang Sehat*. Emc. <https://www.emc.id/id/care-plus/gizi-seimbang-untuk-gaya-hidup-yang-sehat#:~:text=Gizi seimbang adalah susunan makanan,kelamin%2C umur dan status kesehatan>.
- Irfiani, E., & Rani, S. S. (2018). Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 6(4), 165–172. <https://doi.org/10.26418/justin.v6i4.29024>

- Kaggle. (2020). *How to check accuracy for k means algorithm ?* <https://www.kaggle.com/questions-and-answers/175889#977995>
- Nalendra, A. K. (2018). Pengukuran Keakuratan Metode K-Means untuk Menentukan Status Gizi Balita. *JURNAL EKONOMI DAN TEKNIK INFORMATIKA*, 6(2), 48–54.
- Sari, E. A. (n.d.). Penerapan Algoritma K-Means Untuk Menentukan Tingkat Kesehatan Bayi Dan Balita Pada Kabupaten Dan Kota Di Jawa Tengah. *Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro*.
- Sitohang, D. W., & Rikki, A. (2019). Implementasi Algoritma K- Means Clustering untuk Mengelompokkan Data Gizi Balita pada Kecamatan Garoga Tapanuli Utara. *KAKIFIKOM (KUMPULAN ARTIKEL KARYA ILMIAH FAKULTAS ILMU KOMPUTER*, 01(02), 80–92.
- Statistik, B. P. (2018). *Prevalensi balita kekurangan gizi menurut Provinsi di Indonesia (PSG) 2016-2018*. <https://www.bps.go.id/indicator/30/1777/1/prevalensi-balita-kekurangan-gizi-menurut-provinsi-di-indonesia-psg-.html>
- Sutanto, T. (n.d.). *Evaluasi Eksternal Clustering “Pairwise” F- β -Score & NMI: Teori & Aplikasi*. <https://tau-data.id/evaluasi-eksternal/>
- University, B., & Irwansyah, E. (n.d.). *CLUSTERING*. <https://socs.binus.ac.id/2017/03/09/clustering>