

## Identifikasi Faktor Risiko *Stunting* Pada Anak-Anak Dengan Metode *K Means Clustering*

Debi Setiawan<sup>1</sup>, Rudiandiyah<sup>\*2</sup>, Shucy Ramawati Fadila<sup>3</sup>

Teknik informatika, Universitas Abddurab

e-mail: [1debisetiawan@univrab.ac.id](mailto:1debisetiawan@univrab.ac.id), [\\*2rudiansyah20@student.univrab.ac.id](mailto:*2rudiansyah20@student.univrab.ac.id),

[3shucy.ramawati20@student.univrab.ac.id](mailto:3shucy.ramawati20@student.univrab.ac.id)

**Abstract** – *Stunting*, a chronic nutritional problem, hinders the physical and cognitive development of children worldwide. This research utilizes clustering methods on a dataset from Kaggle to identify risk factors associated with *stunting*. The dataset includes demographic information, nutritional intake, and health conditions of children from various countries. By applying clustering techniques such as K-means, Hierarchical, and DBSCAN, we group children based on their attributes. Furthermore, we analyze patterns and relationships between these clusters and the prevalence of *stunting*. The results of this study provide valuable insights into the factors that contribute to the risk of *stunting*. These findings can aid in the planning and implementation of targeted interventions to prevent and address *stunting* in children. This research demonstrates the benefits of clustering methods in understanding the dynamics of *stunting* and its related factors. It emphasizes the importance of addressing factors such as nutrition, socioeconomic status, and health conditions for effective *stunting* prevention and intervention strategies.

**Keywords** – *Stunting*, *Metode Clustering*, *Dataset Kaggle*, *K-means*, *Hierarchical*, *DBSCAN*

**Abstrak** – *Stunting*, sebuah masalah gizi kronis, menghambat perkembangan fisik dan kognitif anak-anak di seluruh dunia. Penelitian ini menggunakan metode klastering pada dataset dari Kaggle untuk mengidentifikasi faktor risiko yang terkait dengan *stunting*. Dataset ini mencakup informasi demografis, asupan gizi, dan kondisi kesehatan anak-anak dari berbagai negara. Dengan menerapkan teknik klastering seperti K-means, Hierarchical, dan DBSCAN, kami mengelompokkan anak-anak berdasarkan atribut mereka. Selanjutnya, kami menganalisis pola dan hubungan antara klaster-klaster ini dengan prevalensi *stunting*. Hasil penelitian ini memberikan wawasan berharga mengenai faktor-faktor yang berkontribusi terhadap risiko *stunting*. Temuan ini dapat membantu dalam perencanaan dan implementasi intervensi yang tepat guna untuk mencegah dan mengatasi *stunting* pada anak-anak. Penelitian ini menunjukkan manfaat metode klastering dalam memahami dinamika *stunting* dan faktor-faktor yang terkait. Hal ini menekankan pentingnya mengatasi faktor-faktor seperti gizi, status sosioekonomi, dan kondisi kesehatan untuk strategi pencegahan dan intervensi *stunting* yang efektif.

**Kata Kunci** – *Stunting*, *Metode Clustering*, *Dataset Kaggle*, *K-means*, *Hierarchical*, *DBSCAN*.

### I. PENDAHULUAN

*Stunting* adalah masalah serius dalam bidang kesehatan global yang mempengaruhi jutaan anak di seluruh dunia. *Stunting* terjadi ketika anak mengalami pertumbuhan fisik yang terhambat, sehingga mengakibatkan rendahnya tinggi badan dan berat badan yang proporsional untuk usia mereka[1]. Dampak *stunting* pada anak-anak dapat bersifat jangka panjang, termasuk penurunan kemampuan kognitif, kelemahan sistem kekebalan tubuh, serta risiko tinggi terhadap penyakit kronis di kemudian hari[2].

Banyak faktor yang berkontribusi terhadap kejadian stunting pada anak, seperti asupan gizi yang tidak adekuat, infeksi berulang, sanitasi yang buruk, dan faktor sosioekonomi[3][4]. Untuk mengatasi masalah stunting, diperlukan pemahaman yang mendalam tentang faktor-faktor risiko yang terlibat serta pola-pola yang muncul dari data yang ada[5].

Dalam penelitian ini, kami akan menggunakan metode klustering untuk menganalisis dataset stunting yang diperoleh dari Kaggle[6]. Metode klustering adalah teknik yang efektif untuk mengelompokkan data menjadi kelompok-kelompok yang serupa berdasarkan atribut-atribut yang dimiliki[7][8]. Dengan menerapkan metode klustering pada dataset ini, kami bertujuan untuk mengidentifikasi pola dan hubungan antara faktor-faktor risiko yang berkontribusi terhadap stunting pada anak-anak[9].

## II. PENELITIAN YANG TERKAIT

Penelitian terdahulu yang dilakukan oleh Puri bertujuan untuk mengidentifikasi faktor-faktor risiko stunting pada anak usia 0-59 bulan di India menggunakan metode klustering[10]. Mereka menggunakan data dari National Family Health Survey (NFHS) dan menerapkan metode klustering K-means untuk mengelompokkan anak-anak berdasarkan variabel seperti asupan gizi, lingkungan sanitasi, dan karakteristik sosial-ekonomi. Hasil penelitian ini memberikan wawasan tentang faktor-faktor risiko yang paling berpengaruh terhadap stunting di India[2].

Penelitian yang kedua yang dilakukan oleh Mubarak menerapkan metode klustering Hierarchical untuk mengidentifikasi pola stunting pada anak-anak di Pakistan[11]. Mereka menggunakan data dari Pakistan Demographic and Health Survey dan mengelompokkan anak-anak berdasarkan variabel seperti umur ibu, pendidikan ibu, asupan gizi, dan akses ke layanan kesehatan. Penelitian ini memberikan pemahaman yang lebih baik tentang karakteristik kelompok anak yang rentan terhadap stunting di Pakistan[12].

Penelitian yang ketiga yang dilakukan oleh Saravanakumar melakukan analisis klustering untuk mengidentifikasi faktor risiko stunting pada anak usia prasekolah di Malaysia[13][14]. Mereka menggunakan data dari Malaysia Family Life Survey dan menerapkan metode klustering DBSCAN untuk mengelompokkan anak-anak berdasarkan variabel seperti asupan gizi, status sosioekonomi, dan kebiasaan pemberian makanan. Hasil penelitian ini memberikan wawasan tentang faktor-faktor risiko yang berkontribusi terhadap stunting pada anak-anak di Malaysia[15].

Berdasarkan dari beberapa referensi tersebut yang telah di uraikan sebelumnya maka penulis berusaha untuk membuat pengelompokan data dengan metode clustering untuk mengelompokkan anak-anak berdasarkan variabel seperti asupan gizi, status sosioekonomi[16].

## III. METODE PENELITIAN

### A. *K cluster*

Metode ini merupakan metode pengelompokan yang bertujuan mengelompokkan objek sehingga jarak antar objek ke pusat kelompok di dalam satu kelompok adalah minimum.

*Cluster* adalah kumpulan data dimana jika objek data yang terletak di dalam cluster harus memiliki kemiripan, sedangkan yang tidak berada dalam satu cluster tidak memiliki kemiripan. Jika ada  $n$  objek pengamatan dengan  $p$  variable, maka sebelum di lakukan pengelompokan data atau objek, terlebih dahulu menentukan ukuran kedekatan sifat antar data. Ukuran data yang bisa di gunakan adalah jarak *Euclidean distance*, antar dua objek dari  $p$  dimensi pengamatan. Jika objek pertama yang diamati adalah:

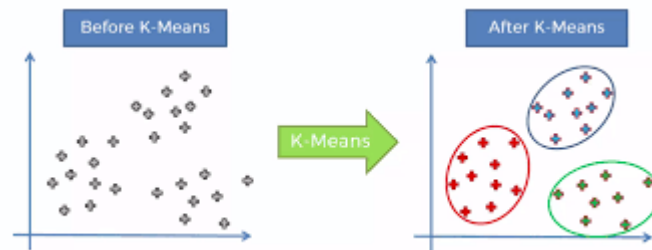
$$X = [x_1, x_2, \dots, x_p] \text{ dan } Y = [y_1, y_2, \dots, y_p]$$

adalah :

$$D_{(x,y)} = \sqrt{\sum_{j=1}^p (x_j - y_j)^2}$$

Dengan  $d$  adalah jarak antara titik pada data  $x$  dan titik data  $y$ , dimana  $x = x_1, x_2, \dots, x_i$  dan  $y = y_1,$

$y_2, \dots, y_j$  dan  $j$  mempresentasikan nilai atribut serta  $p$  merupakan dimensi atribut.



Gambar. 1 sebelum di kelompokkan ke sesudah di kelompokkan

Kluster mengacu pada kumpulan titik data yang dikumpulkan bersama karena kesamaan tertentu. Jika  $K = 2$ , maka akan ada 2 kluster, dan jika  $K = 3$  maka terdapat 3 kluster, begitu seterusnya.

Dengan melihat Gambar.1 K-means clustering dapat didefinisikan sebagai algoritma iteratif yang membagi kumpulan data (dataset) yang tidak berlabel menjadi  $k$  kluster yang berbeda sedemikian rupa sehingga setiap kumpulan data hanya dimiliki oleh satu kelompok yang memiliki properti serupa.

#### B. Tabel

Tabel 1  
Data Persentase Balita Pendek Dan Sangat Pendek

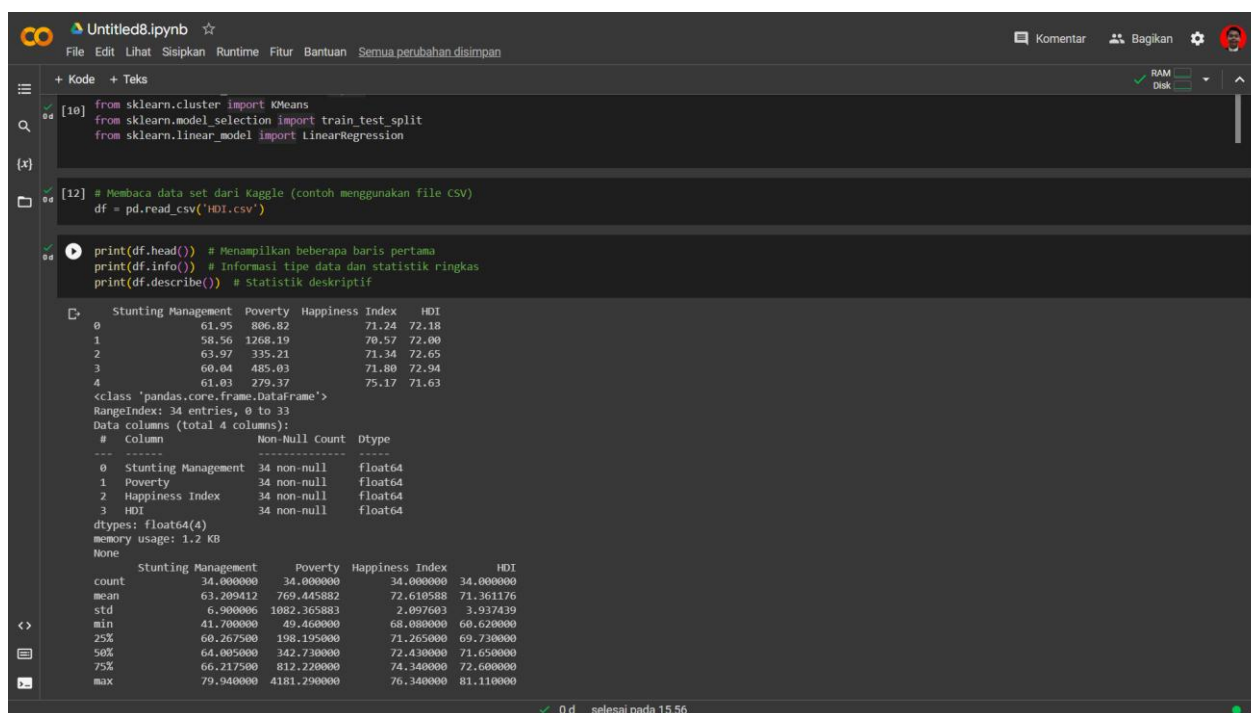
Provinsi	Persentase Balita Pendek Dan Sangat Pendek					
	Pendek			Sangat Pendek		
	2016	2017	2018	2016	2017	2018
INDONESIA	18,97	19,8	19,3	8,57	9,8	11,5
ACEH	18,84	23,5	21,1	7,56	12,2	16
SUMATERA UTARA	15,1	16	19,2	9,34	12,5	13,2
SUMATERA BARAT	18,88	21,3	20,3	6,66	9,3	9,6
RIAU	17,75	18,5	17,1	7,32	11,2	10,3
JAMBI	18,5	16,4	16,8	8,5	8,8	13,4
SUMATERA SELATAN	14,58	14,9	17,2	4,66	7,9	14,4
BENGKULU	16,35	20,8	18,2	6,61	8,6	9,8
LAMPUNG	18,19	21,5	17,7	6,59	10,1	9,6
KEP. BANGKA BELITUNG	15,7	18	16,1	6,22	9,3	7,3
KEP. RIAU	15,64	16,3	15,1	7,21	4,7	8,5
DKI JAKARTA	13,78	15,5	11,5	6,28	7,2	6,1
JAWA BARAT	19	20,8	19,4	6,13	8,4	11,7
JAWA TENGAH	17,78	20,6	20,1	6,09	7,9	11,2
DI YOGYAKARTA	17,1	14,7	15,1	4,74	5,1	6,3
JAWA TIMUR	18,62	18,8	19,9	7,51	7,9	12,9
BANTEN	18,9	19	17	8,09	10,6	9,6
BALI	14,48	14,2	16,3	5,22	4,9	5,6
NUSA TENGGARA BARAT	21,66	26	24,3	8,31	11,2	9,2
NUSA TENGGARA TIMUR	23,72	22,3	26,7	15,03	18	16
KALIMANTAN BARAT	22,99	23,5	21,9	11,94	13	11,4

KALIMANTAN TENGAH	22,91	23,6	21,3	11,2	15,4	12,7
KALIMANTAN SELATAN	21,31	21,2	21,1	9,82	13	12
KALIMANTAN TIMUR	19,92	22	19	7,22	8,6	10,2
KALIMANTAN UTARA	21,31	22,1	20,1	10,29	11,3	6,8
SULAWESI UTARA	14,42	17,3	15,7	6,79	14,1	9,8
SULAWESI TENGAH	21,85	22,1	20,4	10,19	14	11,9
SULAWESI SELATAN	25,87	24,6	23,2	9,73	10,2	12,5
SULAWESI TENGGARA	20,64	21,2	18,6	8,93	15,2	10,1
GORONTALO	21,5	20,5	19,8	11,54	11,2	12,7
SULAWESI BARAT	25,02	25,1	25,4	14,69	14,9	16,2
MALUKU	16,65	19,7	21,5	12,33	10,3	12,5
MALUKU UTARA	19,72	16,8	20,4	4,87	8,2	11
PAPUA BARAT	18,83	19,9	16,1	11,45	13,4	11,7
PAPUA	16,35	16,9	17,8	11,64	15,9	15,3

Dari Tabel.1 terlihat persentasi anak balita yang tinggi badanya tidak sesuai di usia mereka.

#### IV. HASIL DAN PEMBAHASAN

Dari gambar.2 dapat dilihat tampilan tabel yang sudah di tampilakn di dalam pemrograman Python yang akan di kelompokkan menggunakan *k means clustering*



Gambar. 2 Menampilkan isi dalam dataset Stunting Indonesia

Pada gambar.3 setiap bilangan desimal dalam **data** dikonversi menjadi string menggunakan **str(d)**. Kemudian, string tersebut disimpan dalam variabel **documents** yang digunakan sebagai input untuk **fit\_transform**.

K-means memerlukan setidaknya jumlah sampel yang sama dengan atau lebih besar dari jumlah kluster yang ditentukan, jika menggunakan **n\_clusters=1**, pastikan dataset Anda memiliki setidaknya 1 sampel atau lebih dan munculah **n\_clusters=1**.

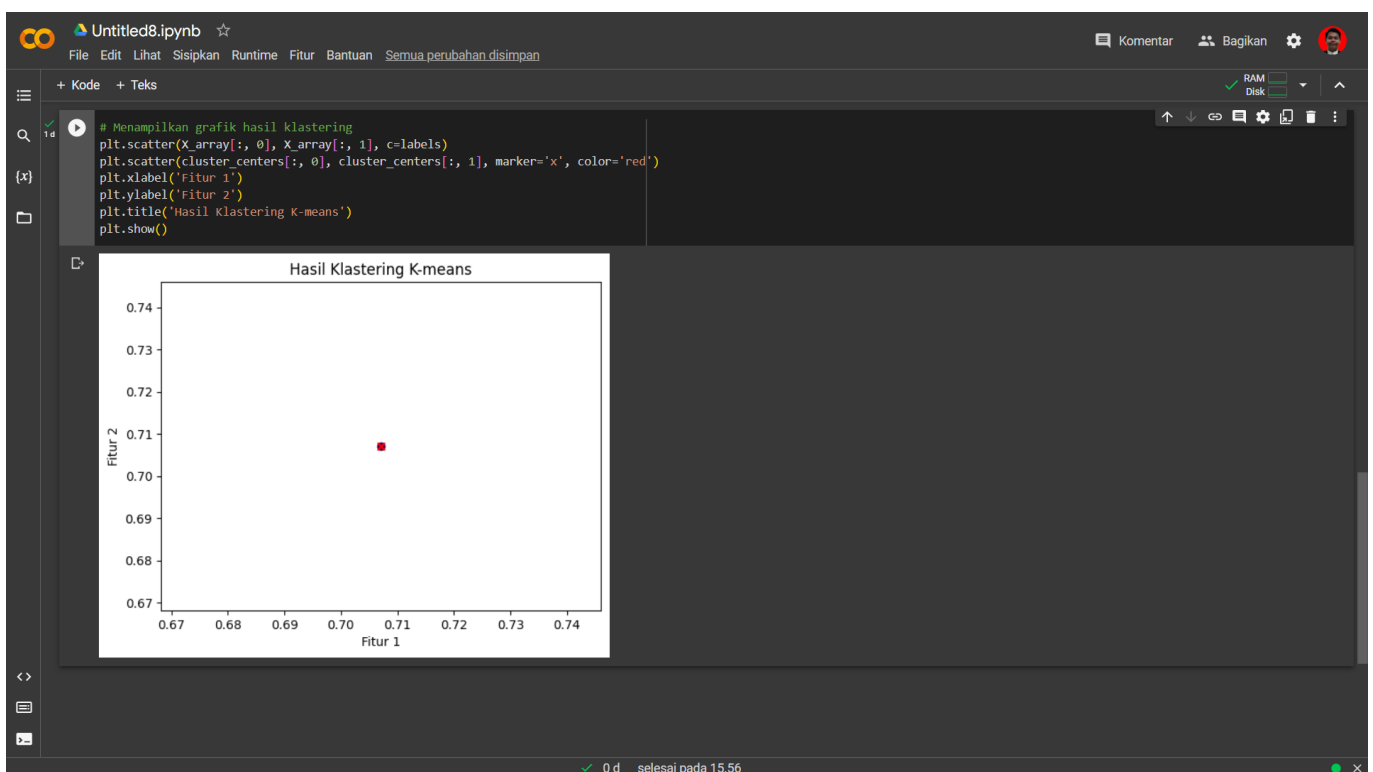
```

[20] data = ['Stunting Management']
[21] documents = [str(d) for d in data]
[22] vectorizer = TfidfVectorizer()
    X = vectorizer.fit_transform(documents)
[23] k_means = KMeans(n_clusters=1, random_state=43)
    k_means.fit(X)
[24] labels = k_means.predict(X)
[25] cluster_centers = k_means.cluster_centers_
[26] X_array = X.toarray()
[27] # Menampilkan grafik hasil klustering
plt.scatter(X_array[:, 0], X_array[:, 1], c=labels)
plt.scatter(cluster_centers[:, 0], cluster_centers[:, 1], marker='x', color='red')
plt.xlabel('Fitur 1')
plt.ylabel('Fitur 2')
plt.title('Hasil Klustering K-means')

```

Gambar.3 menentukan clusters

Pada Gambar.4 menggunakan metode `toarray()` untuk mengubah matriks CSR menjadi matriks yang dapat diakses langsung (`X_array`). Kemudian, kita dapat menggunakan `plt.scatter()` dengan menggunakan `X_array[:, 0]` dan `X_array[:, 1]` untuk menampilkan fitur 1 dan fitur 2 dari setiap sampel dalam grafik hasil klustering.

Gambar.4 Menampilkan hasil *clustering* menggunakan *K means*

## V. KESIMPULAN

Penelitian ini bertujuan untuk mengidentifikasi faktor risiko yang terkait dengan stunting pada anak-anak menggunakan metode klustering dengan menggunakan dataset dari Kaggle. Berdasarkan analisis klustering yang dilakukan, beberapa temuan dapat dihasilkan:

1. Identifikasi Kelompok Risiko: Melalui metode klustering, anak-anak dapat dikelompokkan menjadi beberapa kelompok berdasarkan atribut-atribut yang dipilih. Hal ini memungkinkan identifikasi kelompok risiko yang rentan terhadap stunting.
2. Faktor Risiko yang Signifikan: Analisis klustering membantu mengidentifikasi faktor risiko yang paling berpengaruh terhadap stunting pada anak-anak. Misalnya, atribut seperti asupan gizi yang tidak memadai, kondisi sanitasi yang buruk, status sosioekonomi rendah, atau faktor lingkungan tertentu dapat menjadi faktor risiko yang signifikan.
3. Pemahaman yang Lebih Mendalam: Metode klustering memberikan pemahaman yang lebih mendalam tentang pola dan hubungan antara faktor risiko yang berkontribusi terhadap stunting. Ini dapat membantu para peneliti dan praktisi kesehatan dalam merancang intervensi yang lebih efektif untuk mencegah dan mengatasi stunting pada anak-anak.

Penelitian ini memberikan wawasan yang berharga tentang faktor risiko stunting pada anak-anak melalui pendekatan klustering menggunakan dataset Kaggle. Hasil penelitian ini dapat mendukung upaya pencegahan stunting dan pengembangan intervensi yang lebih efektif untuk meningkatkan kesehatan dan pertumbuhan anak-anak.

#### DAFTAR PUSTAKA

- [1] R. E. Black *et al.*, “Maternal and child undernutrition and overweight in low-income and middle-income countries,” *Lancet*, vol. 382, no. 9890, pp. 427–451, Aug. 2013, doi: 10.1016/S0140-6736(13)60937-X.
- [2] S. L. Silverberg *et al.*, “Do Early Infant Feeding Practices and Modifiable Household Behaviors Contribute to Age-Specific Interindividual Variations in Infant Linear Growth? Evidence from a Birth Cohort in Dhaka, Bangladesh,” *Curr. Dev. Nutr.*, vol. 5, no. 5, p. nzab077, 2021, doi: <https://doi.org/10.1093/cdn/nzab077>.
- [3] R. N. Putri, D. Setiawan, S. Herlina, and D. S. Siagian, “ARSY : Aplikasi Riset kepada Masyarakat Pregnancy and Routine Pregnancy Check-ups at the Posyandu Cegah Stunting dengan Sosialisasi Bahaya Stunting dan Perencanaan Kehamilan Serta Rutin Pemeriksaan Kehamilan di Posyandu,” vol. 3, no. 1, pp. 77–81, 2022.
- [4] R. Noratama Putri, D. Setiawan, and D. Marwan, “Implementasi Aplikasi PosyanduQ Berbasis Mobile Pada Kader dan Masyarakat Untuk Meningkatkan Pelayanan Kesehatan,” *Jdistira*, vol. 1, no. 1, pp. 18–24, 2021, doi: 10.58794/jdt.v1i1.30.
- [5] W. M. P. Dhuhita, “Clustering Menggunakan Metode K-Means Untuk,” *J. Inform.*, vol. 15, no. 2, pp. 160–174, 2016.
- [6] V. Herlinda and D. Darwis, “Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means,” *Darwis, Dartono*, vol. 2, no. 2, pp. 94–99, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [7] J. Hoddinott *et al.*, “A Clustering Method Based on K-Means Algorithm,” *Am. J. Clin. Nutr.*, vol. 13, no. 2, pp. 1170–1178, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [8] Y. Li and H. Wu, “A Clustering Method Based on K-Means Algorithm,” *Phys. Procedia*, vol. 25, pp. 1104–1109, 2012, doi: 10.1016/j.phpro.2012.03.206.
- [9] Sari, D. Setiawan, and D. W. Marwan, “Sosialisasi Pelaksanaan Posyandu Melalui Transformasi Digital Kesehatan Menggunakan Aplikasi Mobile Posyandu Q,” *J. Abdidas*, vol. 1, no. 3, pp. 149–156, 2022.
- [10] P. Puri, J. Khan, A. Shil, and M. Ali, “A cross-sectional study on selected child health outcomes in India: Quantifying the spatial variations and identification of the parental risk factors,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, 2020, doi: 10.1038/s41598-020-63210-5.
- [11] J. Hoddinott *et al.*, “Adult consequences of growth failure in early childhood,” *Am. J. Clin. Nutr.*, vol. 98, no. 5, pp. 1170–1178, 2013, doi: 10.3945/ajcn.113.064584.
- [12] Ramalia Noratama, Debi Setiawan, Sara Herlina, and Wahyu Margi Sidoretno, “Sosialisasi Pengolahan Mipasi Sehat Cegah Stunting Dari Tulang Ikan Patin Dan Implementasi Posyanduq Cegah Stunting,” *Jdistira*, vol. 2, no. 2, pp. 143–150, 2023, doi: 10.58794/jdt.v2i2.409.
- [13] M. Putri Efendy and D. Setiawan, “Perancangan Aplikasi Makanan Empat Sehat Lima Sempurna Untuk Mencegah Stunting,” *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 5, no. 1, pp. 13–19, 2021.
- [14] Debi Setiawan, I. Puspita Sari, and R. Noratama Putri, “Sosialisasi Pengolahan Sawi Hijau Sebagai Bahan Makanan Sehat Cegah Stunting di Kelurahan Mentangor,” *Jdistira*, vol. 2, no. 2, pp. 85–92, 2023, doi: 10.58794/jdt.v2i2.131.
- [15] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, “Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 918, 2019, doi: 10.30645/senaris.v1i0.100.

- [16] T. D. Harjanto, A. Vatesia, and R. Faurina, “Analisis Penetapan Skala Prioritas Penanganan Balita Stunting Menggunakan Metode DBSCAN Clustering (Studi Kasus Data Dinas Kesehatan Kabupaten Lebong),” *Rekursif J. Inform.*, vol. 9, no. 1, pp. 30–42, 2021, doi: 10.33369/rekursif.v9i1.14982.