

Algoritma K-Means untuk *Clustering* Provinsi di Indonesia Berdasarkan Kasus Stunting

Syelvian Julianti^{a1}, I Made Widiartha^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹syelviavia06@gmail.com
²madewidiartha@unud.ac.id

Abstract

Stunting is a nutritional issue that poses a global challenge, especially in developing countries like Indonesia. According to UNICEF, Indonesia ranks among the top five countries with the highest stunting prevalence. To address this issue, clustering provinces in Indonesia each year can help ensure equitable food distribution and other resources. This can be done using the K-Means clustering algorithm, with the optimal number of clusters determined by the elbow method and evaluated using the silhouette coefficient and Davies-Bouldin index. The optimal number of clusters was found to be 3, with a silhouette coefficient of 0.50 and a Davies-Bouldin index of 0.70. In 2020, there were 15 provinces in cluster 1, 6 provinces in cluster 2, and 17 provinces in cluster 3. In 2021, 15 provinces were in cluster 1, 17 in cluster 2, and 6 in cluster 3. In 2022, there were 17 provinces in cluster 1, 14 in cluster 2, and 7 in cluster 3. In 2023, 5 provinces were in cluster 1, 14 in cluster 2, and 19 in cluster 3. By 2024, there were 18 provinces in cluster 1, 17 in cluster 2, and 3 in cluster 3.

Keywords: *Stunting, K-Means, Elbow Method, Silhouette Coefficient, Davies Bouldin Index*

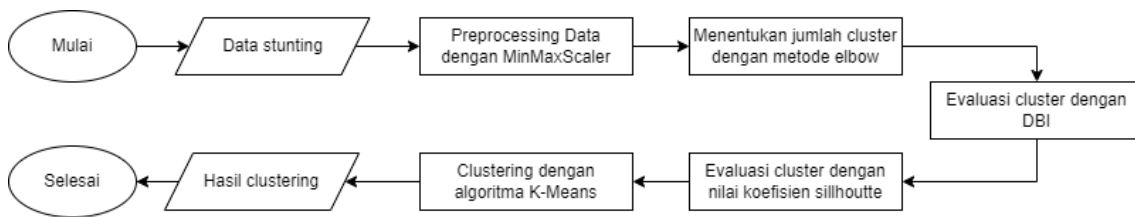
1. Pendahuluan

Salah satu permasalahan gizi yang menjadi permasalahan global khususnya negara berkembang seperti Indonesia adalah stunting. Menurut WHO stunting adalah gangguan kembang tumbuh anak karena terserang infeksi, kurangnya asupan gizi, ataupun stimulasi yang kurang memadai. Menurut UNICEF, Indonesia termasuk dalam lima negara dengan kasus stunting tertinggi dari 88 negara di seluruh dunia yang menjadikannya salah satu negara berkembang dengan prevalensi stunting tinggi. Wakil Ketua Komisi IX DPR RI Kurniasi Mufidayati mengatakan bahwa pada pertengahan tahun 2023 prevalensi stunting di Indonesia mencapai 21,6 persen, sedangkan target yang ingin dicapai pada tahun 2024 adalah 14 persen [1]. Sehingga, untuk mencapai target tersebut Indonesia harus mengurangi kasus stunting sebesar 7,6 persen dalam kurun waktu satu tahun. Untuk mencapai tujuan tersebut, langkah awal yang dapat dilakukan adalah mengelompokkan provinsi berdasarkan persentase kasus stunting sehingga dapat dilakukan pemerataan akses kesehatan, pangan, dan pendidikan gizi. Pengelompokkan dapat dilakukan dengan menggunakan algoritma *clustering* seperti K-Means. Algoritma K-Means digunakan karena algoritma ini memiliki kompleksitas yang rendah sehingga dinilai lebih efisien [2]. Tetapi, algoritma K-Means memiliki kekurangan dalam penentuan jumlah kluster. Oleh karena itu, dibutuhkan metode yang mampu menentukan jumlah cluster salah satunya adalah metode *elbow*. Tetapi, metode *elbow* dianggap kurang efektif karena terkadang bersifat ambigu [3]. Untuk memastikan bahwa jumlah *cluster* dengan metode *elbow* telah optimal maka dilakukan evaluasi dengan *silhouette coefficient* dan *Davies Bouldin Index* (DBI). Dalam penelitian [4] algoritma K-Means digunakan dalam klasifikasi sekolah dasar, didapatkan 5 *cluster* dengan skor *silhouette* kurang dari 0,71 dengan persentase *cluster* 1 21,43%, *cluster* 2 28,25%, *cluster* 3 14,28%, *cluster* 4 21,42%, dan *cluster* 5 14,29%, dari hasil tersebut Pemerintah Kota Pangkalpinang dapat mengklasifikasikan sekolah tingkat dasar dengan baik. Dalam penelitian [5] algoritma K-Means digunakan dalam pengelompokkan daerah berdasarkan jumlah industri, didapatkan bahwa nilai DBI terbaik adalah 0,175 pada saat jumlah *cluster* 3 sehingga didapatkan kesimpulan bahwa

evaluasi nilai DBI pada algoritma K-Means menghasilkan optimasi jumlah *cluster*. Berdasarkan latar belakang permasalahan dan penelitian terdahulu, maka algoritma K-Means akan digunakan untuk *clustering* provinsi di Indonesia berdasarkan kasus stunting dengan menggunakan metode *elbow* untuk penentuan jumlah *cluster*, lalu dievaluasi dengan *silhouette coefficient* dan *Davies Bouldin Index (DBI)*.

2. Metode Penelitian

Gambar 1 merupakan tahapan penelitian yang akan dilakukan. Input yang digunakan berupa persentase kasus stunting seluruh provinsi di Indonesia. *Preprocessing* data dilakukan dengan *MinMaxScaler*. Penentuan jumlah *cluster* menggunakan metode *elbow*, lalu dievaluasi dengan *silhouette coefficient* dan DBI. Selanjutnya dilakukan *clustering* dengan algoritma K-Means sehingga mendapatkan hasil *cluster*.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data yang digunakan pada penelitian ini adalah persentase kasus stunting 38 provinsi di Indonesia mulai dari tahun 2020 hingga 2024 yang didapatkan pada [website aksi.bangda.kemendagri.go.id](https://aksi.bangda.kemendagri.go.id). Atribut yang digunakan adalah provinsi dan persentase kasus stunting setiap tahun dimana *clustering* akan dilakukan pada setiap provinsi dalam satu tahun.

2.2 Preprocessing Data

Dalam penelitian ini *preprocessing* data berupa normalisasi data dengan *MinMaxScaler*. Normalisasi data berguna untuk penskalaan kembali terhadap nilai, sehingga proses pengolahan data menjadi lebih mudah [6]. *MinMaxScaler* merupakan teknik normalisasi data yang mengubah nilai pada data dalam skala nilai 0 dan 1 [7].

2.3 Jumlah Cluster

Dalam melakukan *clustering* penentuan jumlah *cluster* adalah hal yang utama, tetapi pada algoritma K-Means penentuan jumlah *cluster* dilakukan secara *random*. Oleh karena itu, dalam penentuan jumlah *cluster* menggunakan metode *elbow*. Tetapi, metode *elbow* terkadang ambigu dan bersifat subjektif sehingga perlu dilakukan evaluasi menggunakan *silhouette coefficient* dan DBI.

a. Metode *Elbow*

Menurut [8] metode *elbow* adalah metode untuk menentukan jumlah *cluster* berdasarkan persentase perbandingan nilai k atau jumlah *cluster* pada titik tertentu. Hasil dari metode *elbow* direpresentasikan dalam bentuk grafik dengan penentuan jumlah *cluster* terbaik apabila terdapat penurunan nilai terbesar sehingga membentuk sudut pada grafik. Grafik tersebut terbentuk dari hasil perhitungan SSE (Sum of Square Error) yang dibandingkan [9]. SSE pada K-Means dihitung menggunakan rumus pada persamaan 1 [10].

$$SSE = \sum_{k=1}^K \sum_{x_i} |x_i - c_k|^2 \quad (1)$$

Keterangan:

K = Cluster ke-c
 c_k = Pusat data ke-i
 x_i = Jarak data ke-i

b. *Silhouette Coefficient*

Silhouette coefficient mengevaluasi *cluster* dengan melihat kualitas objek pada suatu *cluster*. Nilai *silhouette coefficient* dihitung dari jumlah *cluster* 2 hingga jumlah *cluster* n-1 dengan mencari nilai tertinggi dari *Silhouette Index Global*, rumus *silhouette coefficient* terdapat persamaan 2 [10].

$$SC = maks_k SI(k) \tag{2}$$

Keterangan:

SC = *Silhouette coefficient*
 SI = *Silhouette index global*
 k = Jumlah *cluster*

Hasil perhitungan *silhouette coefficient* berada pada rentang nilai 1 dan -1, dengan hasil perhitungan terbaik ketika mendekati 1 karena mengindikasikan kualitas hasil *clustering* yang baik.

c. *Devies Bouldin Index (DBI)*

DBI merupakan metode untuk mengevaluasi *cluster* berdasarkan ukuran *cluster* dan pesebarannya [11]. Nilai DBI yang baik adalah nilai DBI yang mendekati 0 karena hal tersebut menunjukkan bahwa hasil *cluster* yang diperoleh baik dan optimal. Perhitungan DBI dapat dilakukan dengan menggunakan persamaan 3 [12].

$$DBI = \frac{1}{C} \sum_{i=1}^C R_i \tag{3}$$

$$R_i = \max_{j=1, \dots, C, i \neq j} R_{ij}, R_{ij} = \frac{S_i + S_j}{d(V_i, V_j)} \tag{4}$$

Keterangan:

C = Jumlah *cluster*
 R_{ij} = Ukuran kemiripan *cluster* ke-i dan ke-j
 S_i = Ukuran dispresi *cluster* i

2.4. K-Means

K-Means merupakan algoritma *clustering* yang telah umum digunakan karena mampu melakukan *clustering* dalam jumlah data besar dan mudah diimplementasikan. Algoritma K-Means bertujuan meminimalkan kemiripan data antar *cluster* dan memaksimalkan kesamaan data dalam satu *cluster* [13]. Kemiripan data dihitung berdasarkan jarak data dengan centroid (titik pusat data). Perhitungan jarak data dengan centroid dapat dilakukan dengan rumus pada persamaan 5 yang merupakan rumus *Euclidean Distance*.

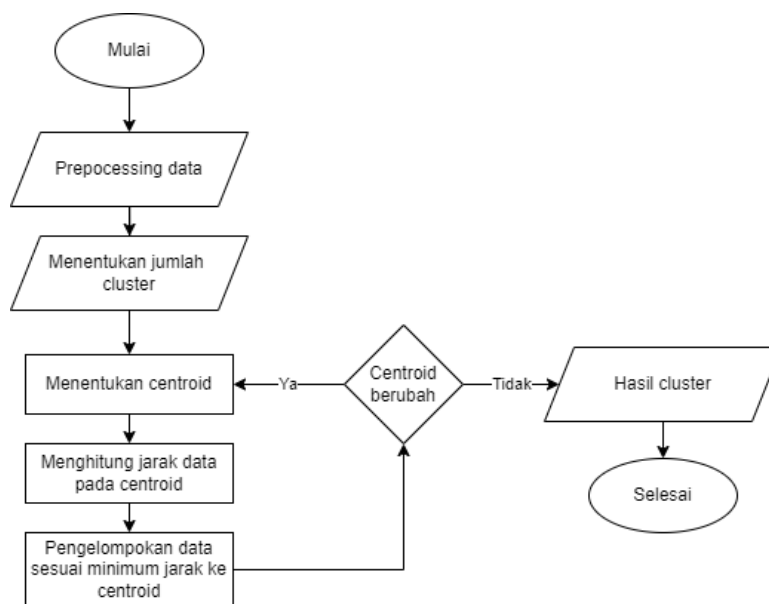
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{5}$$

Keterangan:

$d(x, y)$ = Jarak antar data x ke y
 x_i = Data uji ke-i
 y_i = Data latih ke-i

Flowchart algoritma K-Means terdapat pada gambar 2 dengan detail tahapan yang dilakukan algoritma K-Means, yaitu:

- Melakukan *preprocessing* data berupa normalisasi data atau lainnya yang berguna untuk memaksimalkan hasil *clustering*.
- Menentukan jumlah *clustering* atau pada algoritma K-Means disebut dengan nilai K.
- Menentukan *centroid* atau pusat data sesuai dengan jumlah K yang telah ditentukan.
- Menghitung jarak data dari *centroid*.
- Melakukan pengelompokan data berdasarkan jarak terdekat atau minimum antara *centroid* dengan data.
- Dilakukan pemeriksaan, apabila *centroid* mengalami perpindahan maka akan dilakukan ulang penentuan *centroid* tetapi jika tidak ada perpindahan maka hasil pengelompokan atau *clustering* akan menjadi hasil akhir dari proses *clustering*.



Gambar 2. Flowchart Algoritma K-Means

3. Hasil dan Diskusi

Dalam penelitian ini *clustering* data menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka scikit-learn. Dengan tahapan berupa *preprocessing*, menentukan jumlah *cluster* menggunakan metode *elbow*, evaluasi *cluster* dengan silhouette coefficient dan DBI, lalu *clustering* dengan K-Means.

3.1 Preprocessing Data

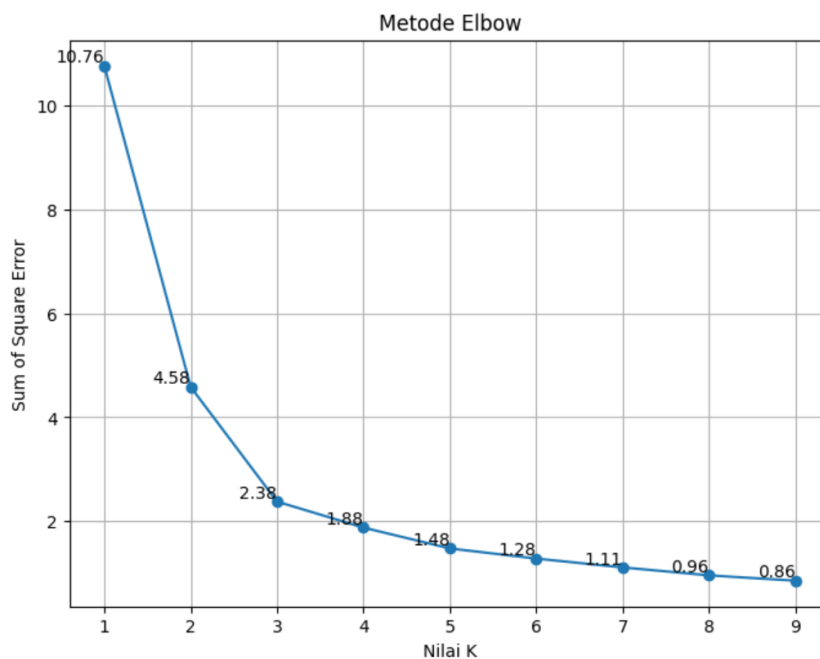
Tahapan *preprocessing* data yang dilakukan berupa normalisasi dengan menggunakan metode MinMaxScaler. Gambar 3 adalah data hasil normalisasi pada tahun 2020 dengan data yang diurutkan berdasarkan nama provinsi yang terurut secara menyamping sesuai pada [website aksi.bangda.kemendagri.go.id](http://website.aksi.bangda.kemendagri.go.id). Dengan menggunakan metode MinMaxScaler maka data akan berada pada skala nilai 0 dan 1.

Hasil Normalisasi Data 2020:
 [0.46503497 0.23426573 0.60839161 0.26573427 0.26223776 0.07692308
 0.28671329 0.18181818 0.11188811 0.31818182 0. 0.31118881
 0.45454545 0.41258741 0.3986014 0.16783217 0.32517483 0.7027972
 0.91608392 0.95454545 0.77272727 0.45454545 0.5034965 1.
 0.18181818 0.47552448 0.41258741 0.58041958 0.38111888 0.79020979
 0.28671329 0.32167832 0.41608392 0.16083916 0.2027972 0.39160839
 0.57692308 0.61538462]

Gambar 3. Data Tahun 2020 Setelah Normalisasi

3.2 Penentuan Jumlah Cluster

Sebelum melakukan *clustering* perlu ditetapkan nilai *k* atau jumlah *cluster* menggunakan metode *elbow*. Metode *elbow* direpresentasikan dalam bentuk grafik berdasarkan nilai *Sum of Square Error* (SSE). Penentuan jumlah *cluster* ditentukan berdasarkan nilai SSE yang memiliki penurunan signifikan, pada gambar 4 dapat dilihat ketika nilai *k* = 3 terdapat penurunan signifikan. Sehingga, berdasarkan metode *elbow* dapat ditentukan jumlah *cluster* yang digunakan adalah 3.



Gambar 4. Metode Elbow

3.3 Evaluasi Cluster

Setelah ditentukan bahwa jumlah *cluster* adalah 3 maka dilakukan evaluasi menggunakan *silhouette coefficient* dan DBI. *Silhouette coefficient* digunakan untuk mengevaluasi kemiripan satu data dengan data lainnya dalam suatu *cluster*, lalu DBI digunakan untuk mengevaluasi pemisahan antar *cluster*. Nilai *silhouette coefficient* yang terbaik adalah yang mendekati 1, sedangkan nilai DBI yang terbaik adalah yang mendekati 0. Pada tabel 1 dapat dilihat bahwa hasil *silhouette coefficient* terbaik terdapat pada nilai *cluster* 3 dengan nilai 0,50 dan nilai DBI terbaik juga didapatkan pada saat jumlah *cluster* 3 dengan nilai 0,70 yang dapat dilihat pada tabel 2. Dari kedua hasil evaluasi tersebut dan metode *elbow* yang dilakukan sebelumnya didapatkan bahwa jumlah *cluster* yang optimal adalah 3.

Tabel 1. Silhoutte Score

Cluster	Silhouette Score
2	0,47
3	0.50
4	0.39
5	0.39
6	0.38
7	0.28
8	0.28
9	0.27

Tabel 2. Davies-Bouldin Index

Cluster	Davies-Bouldin Index
2	0,80
3	0.70
4	0.90
5	0.89
6	0.83
7	0.77
8	0.88
9	0.76

3.4 Clustering Data

Setelah mendapatkan jumlah *cluster* yang optimal maka akan dilakukan *clustering* data menggunakan algoritma K-Means. *Clustering* dilakukan berdasarkan kasus stunting pada setiap provinsi di Indonesia setiap tahunnya dari tahun 2020 hingga 2024. Karena jumlah *cluster* yang optimal adalah tiga maka didapatkan hasil *cluster* 0, *cluster* 1, dan *cluster* 2. Hasil *cluster* pada tahun 2020 terdapat pada tabel 3 dimana terdapat 15 provinsi pada *cluster* 1, 6 provinsi pada *cluster* 2, dan 17 provinsi pada *cluster* 3.

Tabel 3. Hasil *Clustering* Tahun 2020

Cluster 1	Cluster 2	Cluster 3
Aceh	Nusa Tenggara Barat	Sumatera Utara
Sumatera Barat	Nusa Tenggara Timur	Riau
Jawa Tengah	Kalimantan Barat	Jambi
Di Yogyakarta	Kalimantan Tengah	Sumatera Selatan
Jawa Timur	Kalimantan Utara	Bengkulu
Kalimantan Selatan	Sulawesi Barat	Lampung

Cluster 1	Cluster 2	Cluster 3
Kalimantan Timur		Kepulauan Bangka Belitung
Sulawesi Tengah		Kepulauan Riau
Sulawesi Selatan		Dki Jakarta
Sulawesi Tenggara		Jawa Barat
Gorontalo		Banten
Papua Barat		Bali
Papua		Sulawesi Utara
Papua Tengah		Sulawesi Utara
Papua Barat Daya		Maluku
		Maluku Utara
		Sulawesi Utara

Hasil *cluster* pada tahun 2021 terdapat pada tabel 4 dimana terdapat 15 provinsi pada *cluster* 1, 17 provinsi pada *cluster* 2, dan 6 provinsi pada *cluster* 3.

Tabel 4. Hasil *Clustering* Tahun 2021

Cluster 1	Cluster 2	Cluster 3
Aceh	Sumatera Utara	Nusa Tenggara Barat
Sumatera Barat	Riau	Nusa Tenggara Timur
Jawa Tengah	Jambi	Kalimantan Barat
Di Yogyakarta	Sumatera Selatan	Kalimantan Utara
Jawa Timur	Bengkulu	Sulawesi Tenggara
Kalimantan Tengah	Lampung	Sulawesi Barat
Kalimantan Selatan	Kepulauan Bangka Belitung	
Kalimantan Timur	Kepulauan Riau	
Sulawesi Tengah	Dki Jakarta	
Sulawesi Selatan	Jawa Barat	
Maluku Utara	Banten	
Papua Barat	Bali	
Papua	Sulawesi Utara	
Papua Selatan	Gorontalo	
Papua Barat Daya	Maluku	
	Papua Pegunungan	
	Papua Tengah	

Hasil *cluster* pada tahun 2022 terdapat pada tabel 5 dimana terdapat 17 provinsi pada *cluster* 1, 14 provinsi pada *cluster* 2, dan 7 provinsi pada *cluster* 3.

Tabel 5. Hasil *Clustering* Tahun 2022

Cluster 1	Cluster 2	Cluster 3
Aceh	Sumatera Utara	Nusa Tenggara Barat
Sumatera Barat	Riau	Nusa Tenggara Timur
Jawa Tengah	Jambi	Kalimantan Barat
Di Yogyakarta	Sumatera Utara	Nusa Tenggara Barat
Jawa Timur	Riau	Nusa Tenggara Timur
Kalimantan Tengah	Jambi	Kalimantan Barat
Kalimantan Selatan	Sumatera Selatan	Kalimantan Utara
Kalimantan Timur	Bengkulu	Sulawesi Utara
Sulawesi Tengah	Lampung	Sulawesi Tenggara
Sulawesi Selatan	Kepulauan Bangka Belitung	Sulawesi Barat

Cluster 1	Cluster 2	Cluster 3
Sulawesi Tenggara	Kepulauan Riau	
Gorontalo	Dki Jakarta	
Maluku	Jawa Barat	
Maluku Utara	Banten	
Papua Barat	Bali	
Papua	Sulawesi Utara	
Papua Tengah	Papua Pegunungan	
Papua Barat Daya		

Hasil *cluster* pada tahun 2023 terdapat pada tabel 6 dimana terdapat 5 provinsi pada *cluster* 1, 14 provinsi pada *cluster* 2, dan 19 provinsi pada *cluster* 3.

Tabel 6. Hasil *Clustering* Tahun 2023

Cluster 1	Cluster 2	Cluster 3
Nusa Tenggara Barat	Sumatera Utara	Aceh
Nusa Tenggara Timur	Riau	Sumatera Barat
Kalimantan Barat	Jambi	Jawa Barat
Sulawesi Barat	Sumatera Selatan	Di Yogyakarta
Papua Selatan	Bengkulu	Jawa Tengah
	Lampung	Jawa Timur
	Kepulauan Bangka Belitung	Kalimantan Tengah
	Kepulauan Riau	Kalimantan Selatan
	Dki Jakarta	Kalimantan Timur
	Banten	Kalimantan Utara
	Bali	Sulawesi Selatan
	Sulawesi Utara	Sulawesi Tenggara
	Sulawesi Tengah	Kalimantan Barat Daya
	Gorontalo	Maluku Utara
	Maluku	Papua Barat
	Papua Pegunungan	Papua
		Papua Tengah

Hasil *cluster* pada tahun 2024 terdapat pada tabel 7 dimana terdapat 18 provinsi pada *cluster* 1, 17 provinsi pada *cluster* 2, dan 3 provinsi pada *cluster* 3.

Tabel 7. Hasil *Clustering* Tahun 2024

Cluster 1	Cluster 2	Cluster 3
Aceh	Sumatera Barat	Nusa Tenggara Timur
Sumatera Utara	Jawa Tengah	Sulawesi Barat
Riau	DI Yogyakarta	Papua Barat Daya
Jambi	Nusa Tenggara Barat	
Sumatera Selatan	Kalimantan Barat	
Bengkulu	Kalimantan Tengah	
Lampung	Kalimantan Selatan	
Kepulauan Bangka Belitung	Kalimantan Timur	
Kepulauan Riau	Kalimantan Utara	
DKI Jakarta	Sulawesi Tengah	
Jawa Barat	Sulawesi Selatan	
Jawa Timur	Sulawesi Tenggara	
Banten	Maluku Utara	

Cluster 1	Cluster 2	Cluster 3
Bali	Papua Barat	
Sulawesi Utara	Papua	
Gorontalo	Papua Selatan	
Maluku	Papua Tengah	
Papua Pegunungan		

4. Kesimpulan

Pada penelitian ini dilakukan *clustering* data mengenai kasus stunting berdasarkan provinsi di Indonesia pada tahun 2020 hingga 2024 didapatkan hasil bahwa data stunting di Indonesia dapat dibagi menjadi 3 *cluster*, yaitu *cluster 1*, *cluster 2*, dan *cluster 3*. Jumlah *cluster* tersebut didapatkan dari pencarian nilai K atau jumlah *cluster* optimal menggunakan metode elbow lalu dievaluasi dengan *silhouette coefficient* dan *davies bouldin index*. Hasil terbaik *silhouette coefficient* dengan nilai 0,50 dan *davies bouldin index* dengan nilai 0,70 pada jumlah *cluster* 3. Oleh karena itu, jumlah *cluster* 3 digunakan untuk *clustering* data stunting setiap tahun di Indonesia. Didapatkan hasil bahwa pada tahun 2020 terdapat terdapat 15 provinsi pada *cluster 1*, 6 provinsi pada *cluster 2*, dan 17 provinsi pada *cluster 3*. Pada tahun 2021 terdapat 15 provinsi pada *cluster 1*, 17 provinsi pada *cluster 2*, dan 6 provinsi pada *cluster 3*. Pada tahun 2022 terdapat terdapat 17 provinsi pada *cluster 1*, 14 provinsi pada *cluster 2*, dan 7 provinsi pada *cluster 3*. Tahun 2023 terdapat 5 provinsi pada *cluster 1*, 14 provinsi pada *cluster 2*, dan 19 provinsi pada *cluster 3*. Lalu, pada tahun 2024 terdapat 18 provinsi pada *cluster 1*, 17 provinsi pada *cluster 2*, dan 3 provinsi pada *cluster 3*.

Daftar Pustaka

- [1] K. Mufidayati, "Kurniasih: Di Tahun 2024, Penurunan Stunting Harus Jadi Fokus Bersama," drp.go.id. Accessed: May 05, 2024. [Online]. Available: <https://www.dpr.go.id/berita/detail/id/48290/t/Kurniasih:%20Di%20Tahun%202024,%20Penurunan%20Stunting%20Harus%20Jadi%20Fokus%20Bersama#:~:text=Dikatakan%20Kurniasih%20pada%20pertengahan%20tahun,adalah%2014%20persen%20pada%202024.>
- [2] H. Mutiasari, T. W. Purboyo, and R. A. Nugrahaeni, "Sistem Rekomendasi Film Menggunakan Metode K-means Clustering," in *eProceedings of Engineering*, 2021, pp. 6755–6764.
- [3] S. Mutalib, N. H. Hasbullah, S. Abdul-Rahman, M. R. Shamsuddin, and A. M. A. Malik, "Herbal Plant Analysis Based on Leaf Features using K-Means Clustering," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, May 2022. doi: 10.1088/1755-1315/1019/1/012026.
- [4] Y. Bagus Pratama and A. Setiawan, "Implementasi Machine Learning Menggunakan Algoritma K-Means Untuk Klasifikasi Sekolah Dasar," *RESOLUSI: Rekayasa Teknik Informatika dan Informasi*, vol. 4, no. 3, pp. 249–257, 2024, [Online]. Available: <https://djournals.com/resolusi>
- [5] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, 2021, [Online]. Available: www.bps.go.id
- [6] R. G. Whendasmoro and Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, pp. 872–876, Aug. 2022, doi: 10.30865/jurikom. v9i4.4526.
- [7] W. Purba *et al.*, "Penerapan Data Mining Untuk Pengelolaan Data Rekam Medis Menggunakan Metode K-means Clustering Pada Rumah Sakit Royal Prima Medan," *Jurnal TEKINKOM*, vol. 6, no. 1, pp. 158–168, 2023, doi: 10.37600/tekinkom. v6i1.857.
- [8] N. Putu, E. Merliana, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," in *Pros. Semin. Nas. MULTI DISIPLIN ILMU CALL*, 2015, pp. 978–979.

- [9] E. Muningsih and S. Kiswati, "Sisitem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan," *JOUTICA*, vol. 3, no. 1, 2018.
- [10] D. Ayu, I. C. Dewi, and K. Pramita, "Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *JURNAL MATRIX*, vol. 9, no. 3, pp. 102–109, 2019.
- [11] A. Putri Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 1, pp. 164–172, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
- [12] I. Rosaliyah, B. Nurhakim, and J. Manajemen, "Clustering Kejadian Bencana Alam di Jawa Barat Berdasarkan Jenis Bencana Menggunakan K-means," *Jurnal Teknik Elektro dan Informatika*, vol. 18, no. 1, pp. 10–16, 2023, [Online]. Available: <https://opendata.jabarprov.go.id/>
- [13] I. Nyoman and M. Adiputra, "Clustering Penyakit DBD Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means," *INSERT: Information System and Emerging Technology Journal*, vol. 2, no. 2, pp. 99–105, 2021.