
**PENANGANAN DATA OUTLIER MENGGUNAAN ROBUST CLUSTER
DALAM PEMETAAN DATASET STUNTING****Kharul Abdi Sinuraya¹, Andrinata², Nurul Azzuma³****Sistem Informasi, Universitas Royal, Kisaran**e-mail: ¹abdikhairul26@gmail.com

Abstract: *The urgency of this research lies in the importance of accurate stunting data mapping to understand the patterns and distribution of this issue within a specific region. However, in data collection and analysis, outliers—values far beyond the normal range—are often present and may affect the validity of the analysis results. Conventional clustering methods, such as K-Means, are highly sensitive to outliers, which may lead to inaccurate and misleading mappings. Therefore, a more robust approach is required to handle outliers, one of which is the robust clustering method. Robust Clustering is an effective technique in dealing with outliers, enabling the identification of homogeneous data groups without being affected by extreme values. This study aims to explore and address the issue of outliers in stunting dataset mapping. Based on the clustering analysis using the K-Means method, three groups of Public Health Centers (Puskesmas) in Asahan Regency were identified based on stunting cases from 2022 to 2024. After outlier detection and handling, the Davies-Bouldin Index decreased from 0.529 to 0.483, indicating improved cluster quality. The final results show that the majority of Public Health Centers fall into Cluster 1 with relatively low to moderate stunting cases, Cluster 2 consists of centers with significantly increasing cases, and Cluster 3 represents those with very high numbers of cases.*

Keywords: *Stunting; Data Outlier; Robust Clustering; Mapping; Machine Learning*

Abstrak: Urgensi penelitian ini terletak pada pemetaan data stunting yang akurat sangat penting untuk memahami pola dan distribusi masalah ini di wilayah tertentu. Akan tetapi, dalam pengumpulan dan analisis data, seringkali terdapat data outlier, nilai yang jauh di luar rentang normal, yang dapat mempengaruhi hasil analisis. Metode pengelompokan konvensional, seperti K-means, sering kali sensitif terhadap outlier, yang dapat menghasilkan pemetaan yang tidak akurat dan menyesatkan. Oleh karena itu, diperlukan pendekatan yang lebih robust untuk menangani data outlier, salah satunya dengan menggunakan metode robust cluster. Metode Robust Clustering merupakan metode pengelompokan yang efektif dalam menghadapi outlier, memungkinkan identifikasi kelompok data yang homogen tanpa dipengaruhi oleh nilai ekstrem. Penelitian ini bertujuan untuk menggali dan mengatasi masalah data outlier dalam pemetaan dataset stunting. Berdasarkan hasil analisis clustering dengan metode K-Means, diperoleh tiga kelompok Puskesmas di Kabupaten Asahan berdasarkan jumlah kasus stunting tahun 2022–2024. Setelah dilakukan deteksi dan penanganan outlier, nilai Davies-Bouldin Index menurun dari 0,529 menjadi 0,483, menunjukkan kualitas cluster yang lebih baik. Hasil akhir menunjukkan bahwa mayoritas Puskesmas masuk ke dalam Cluster 1 dengan kasus stunting relatif rendah hingga sedang, sedangkan Cluster 2 terdiri dari Puskesmas dengan tren kasus meningkat cukup signifikan, dan Cluster 2 berisi Puskesmas dengan jumlah kasus sangat tinggi.

Kata kunci: Stunting; Data Outlier; Robust Clustering; Pemetaan; Machine Learning

PENDAHULUAN

Stunting adalah masalah kesehatan yang signifikan, di mana anak mengalami pertumbuhan terhambat akibat kekurangan gizi dalam waktu lama(1). Di Indonesia, prevalensi stunting masih menjadi tantangan besar, dengan data menunjukkan bahwa sekitar 24,4% anak di bawah lima tahun mengalami stunting pada tahun terakhir. Di Kabupaten Asahan, angka ini bahkan lebih tinggi, mencerminkan kondisi sosial ekonomi dan akses terhadap layanan kesehatan yang masih perlu diperbaiki. Data dari berbagai sumber menunjukkan bahwa stunting dapat mengakibatkan berbagai masalah, seperti penurunan kemampuan kognitif, produktivitas, dan kualitas hidup anak.

Urgensi penelitian ini terletak pada pemetaan data stunting yang akurat sangat penting untuk memahami pola dan distribusi masalah ini di wilayah tertentu. Akan tetapi, dalam pengumpulan dan analisis data, seringkali terdapat data outlier, nilai yang jauh di luar rentang normal, yang dapat mempengaruhi hasil analisis. Data outlier ini bisa disebabkan oleh berbagai faktor, seperti kesalahan pengukuran atau anomali dalam data. Dalam konteks machine learning, outlier dapat mengganggu model prediksi dan analisis yang dibangun, sehingga menghasilkan informasi yang tidak akurat(2,3). Machine learning merupakan cabang kecerdasan buatan yang memungkinkan sistem belajar dari data dan meningkatkan kinerja tanpa pemrograman eksplisit(4).

Metode pengelompokan konvensional, seperti K-means, sering kali sensitif terhadap outlier, yang dapat menghasilkan pemetaan yang tidak akurat dan menyesatkan. Oleh karena itu, diperlukan pendekatan yang lebih robust untuk menangani data outlier, salah satunya dengan menggunakan metode robust cluster. Metode Robust Clustering merupakan metode pengelompokan yang efektif dalam menghadapi outlier,

memungkinkan identifikasi kelompok data yang homogen tanpa dipengaruhi oleh nilai ekstrem(5). Metode robust cluster ini dirancang untuk tetap efektif dalam menghadapi keberadaan outlier, sehingga hasil pemetaan yang dihasilkan dapat lebih dipercaya dan berguna bagi pengambilan keputusan.

Penelitian ini bertujuan untuk menggali dan mengatasi masalah data outlier dalam pemetaan dataset stunting di Kabupaten Asahan, sehingga diharapkan dapat memberikan informasi yang lebih akurat untuk perencanaan intervensi yang tepat dalam mengatasi masalah stunting.

METODE

Prosedur Penelitian

Berikut adalah prosedur penelitian sesuai dengan diagram alir gambar 2.

1. Studi Pendahuluan

Identifikasi Masalah: Memahami isu stunting di Kabupaten Asahan, termasuk banyaknya kasus dan dampaknya terhadap kesehatan anak. Langkah ini penting untuk menetapkan fokus penelitian.

2. Pengumpulan Data

Mengumpulkan data dari berbagai sumber yang relevan, termasuk Dinas Kesehatan, Badan Pusat Statistik (BPS), Puskesmas, dan wawancara dengan orang tua. Data yang komprehensif diperlukan untuk analisis yang akurat.

3. Pra-pemrosesan Data

Membersihkan data dengan menghapus duplikat, menangani nilai yang hilang (missing values), dan memastikan format data konsisten. Identifikasi outlier juga dilakukan untuk mempersiapkan data agar siap untuk analisis lebih lanjut menggunakan normalisasi Z-Score.

4. Fitur Selection

Menggunakan teknik analisis untuk memilih fitur-fitur yang paling berpengaruh terhadap stunting. Ini bisa dilakukan melalui analisis

korelasi atau metode pemilihan fitur berbasis model, seperti Random Forest, untuk meningkatkan akurasi model.

5. Analisis Eksploratori

Melakukan analisis awal untuk memahami distribusi dan karakteristik data. Visualisasi data, seperti histogram atau box plot, digunakan untuk menemukan pola atau anomali yang ada dalam dataset.

6. Pengembangan Model Clustering

Memilih algoritma machine learning yang tepat, K-Medoids dan DBSCAN, untuk mengelompokkan data. Implementasi robust clustering sangat penting untuk menangani outlier dan memastikan hasil yang lebih akurat.

a. K-Medoids

Definisi: K-Medoids adalah algoritma pengelompokan yang mirip dengan K-Means, tetapi menggunakan medoid (titik data yang paling representatif dalam cluster) daripada mean.

Langkah-langkah:

1. Inisialisasi: Pilih K medoid secara acak dari dataset.
2. Penugasan: Setiap titik data ditugaskan ke medoid terdekat.
3. Pembaharuan Medoid: Untuk setiap cluster, pilih titik data yang meminimalkan total jarak intra-cluster (misalnya, menggunakan jarak Manhattan atau Euclidean).
4. Iterasi: Ulangi langkah 2-3 hingga konvergensi (tidak ada perubahan dalam medoid).

b. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Definisi: DBSCAN adalah algoritma pengelompokan berbasis kepadatan yang mengelompokkan titik data

berdasarkan kepadatan titik di sekitarnya. Parameter:

1. ϵ (epsilon): Jarak maksimum untuk dianggap sebagai tetangga.
2. MinPts: Jumlah titik minimum dalam radius ϵ untuk membentuk cluster.

Langkah-langkah:

1. Tentukan Titik Inti: Titik data yang memiliki setidaknya MinPts tetangga dalam jarak ϵ dianggap sebagai titik inti.
2. Buat Cluster: Untuk setiap titik inti, semua titik yang dapat dijangkau (reachable) ditambahkan ke cluster.
3. Tandai Noise: Titik yang tidak termasuk dalam cluster mana pun dianggap sebagai noise.

c. Robust Clustering

Robust clustering mengacu pada teknik pengelompokan yang tahan terhadap outlier dan noise dalam data. Salah satu pendekatan yang umum adalah menggunakan medoid atau teknik berbasis kepadatan.

Pendekatan:

1. Penggunaan Medoid: Seperti K-Medoids, yang lebih tahan terhadap outlier dibandingkan K-Means.
2. Penggunaan Jarak Mahalanobis: Untuk mengukur jarak yang dapat menangani varians dalam data.
3. Metode Berbasis Kepadatan: Seperti DBSCAN, yang mengidentifikasi cluster berdasarkan kepadatan dan dapat mengabaikan noise.

7. Pemetaan Hasil

Menerapkan model clustering yang telah dikembangkan untuk menghasilkan peta stunting. Pemetaan ini akan menunjukkan distribusi stunting di Kabupaten Asahan berdasarkan hasil analisis.

8. Evaluasi Model

Menguji akurasi model dengan menggunakan data tambahan dan melakukan validasi hasil. Jika hasilnya tidak memuaskan, penyesuaian pada model dan teknik yang digunakan akan dilakukan untuk meningkatkan kinerja.

HASIL DAN PEMBAHASAN

Dataset ini berisi data jumlah kasus stunting yang tercatat di berbagai Puskesmas Kabupaten Asahan selama tiga tahun berturut-turut, yaitu tahun 2022, 2023, dan 2024. Data ini disusun dalam bentuk tabel yang memuat informasi sebagai berikut:

1. No: Nomor urut setiap data entri, sebagai pengenalan unik untuk masing-masing Puskesmas.
2. Puskesmas: Nama atau kode identifikasi dari masing-masing pusat kesehatan masyarakat di Kabupaten Asahan. Nama ini digunakan untuk membedakan lokasi dan unit layanan kesehatan yang melaporkan data.
3. Tahun2022: Jumlah kasus stunting yang dilaporkan di setiap Puskesmas selama tahun 2022. Data ini mencerminkan kondisi awal sebelum intervensi jangka menengah dan panjang dilakukan.

4. Tahun2023: Jumlah kasus stunting di setiap Puskesmas selama tahun 2023. Data ini digunakan untuk melihat tren kenaikan atau penurunan kasus setelah adanya program intervensi.

5. Tahun2024: Jumlah kasus stunting di masing-masing Puskesmas selama tahun 2024. Data ini memberi gambaran perkembangan kasus stunting selama tiga tahun terakhir dan menilai efektivitas kebijakan atau program kesehatan yang diimplementasikan.

Karakteristik Dataset:

1. Sifat Data: Kuantitatif dan numerik, merepresentasikan jumlah kasus.
2. Jumlah Data: Terdiri dari 24 entri, masing-masing mewakili Puskesmas berbeda.
3. Fluktuasi Data: Terjadi variasi jumlah kasus dari satu tahun ke tahun berikutnya, menunjukkan tren naik, turun, atau stabil yang dapat dianalisis secara statistik.

Langkah Selanjutnya yaitu melakukan proses analisis data dengan dimulai dari transformasi data menggunakan Z-Score.

Dihasilkan dataset normalisasi

Data setelah transformasi Z-Score:

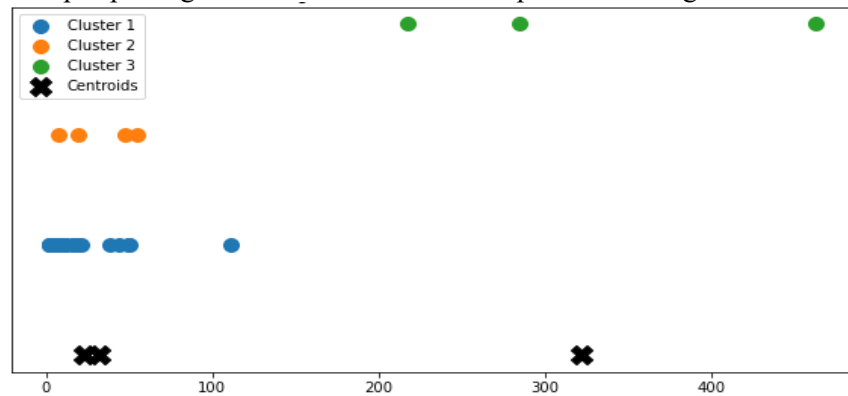
```
[ [ 1.48111216  1.24458581  0.83015845 ]
  [-0.51451151 -0.61278318 -0.8607869 ]
  [-0.51451151 -0.61278318 -0.87673921 ]
  [-0.37264252 -0.47905261 -0.65340681 ]
  [ 0.47857136 -0.37503995  0.43135059 ]
  [-0.47667978 -0.62764213 -0.79697764 ]
  [-0.55234323 -0.67221898 -0.8607869 ]
  [-0.55234323 -0.67221898 -0.87673921 ]
  [-0.55234323 -0.65736003 -0.71721606 ]
  [-0.41993218 -0.67221898 -0.60554986 ]
  [-0.05107283  0.44220241  1.57991724 ]
  [-0.53342737 -0.65736003 -0.82888227 ]
  [ 2.12425155  2.04696921  0.41539827 ]
  [-0.12673629 -0.16701462  2.13824825 ]
  [-0.40101632  1.18515      -0.12698042 ]
  [-0.10782043 -0.55334737 -0.47793134 ]
```

```

[-0.50505357  0.11530547  2.32967602]
[-0.40101632 -0.67221898 -0.47793134]
[-0.15511009 -0.33046309 -0.36626514]
[-0.21185768 -0.0184251  -0.4300744 ]
[-0.49559564 -0.53848842 -0.68531143]
[-0.09836249  0.23417708 -0.01531422]
[-0.45776391 -0.43447575 -0.71721606]
[-0.39155839 -0.048143   0.81420614]
[ 3.80776346  3.5328644  1.83515427]]

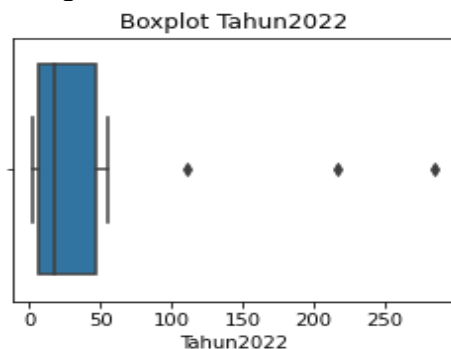
```

Proses selanjutnya meladidkukan anlisis k means tanpa penangana deteksi dan penanganan outlier dengan 3 Cluster. Didapati hasil sebagai berikut:

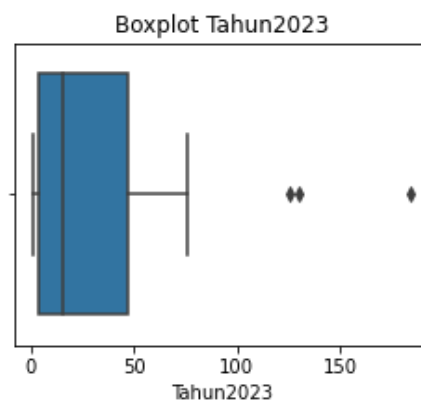


Gambar 1. Clustering tanpa penanganan outlier

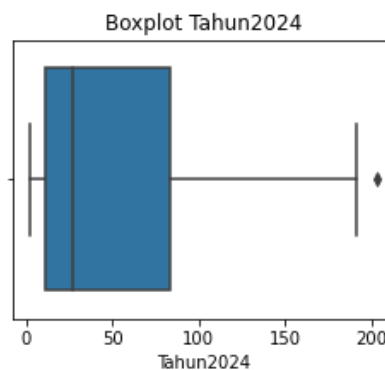
Dengan Davies-Bouldin Index: 0.5293715978060519. Selajutnya dilakukan proses deteksi outlier dan penanganan data outlier.



Gambar 2. Outlier Tahun 2022

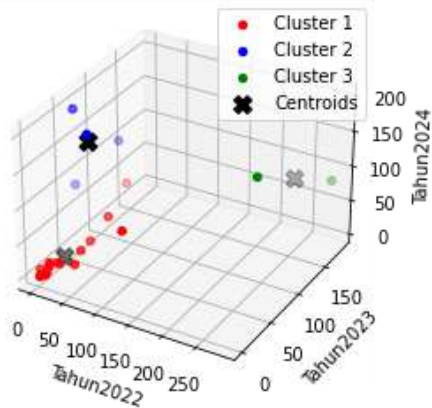


Gambar 3. Outlier Tahun 2023



Gambar 4. Outlier Tahun 2022

Setelah menghapus data outlier, maka didapatkan cluster baru dengan DBI yang lebih baik dari proses berikut:



Gambar 5. Clustering tanpa outlier Dengan Davies-Bouldin Index: 0.4825655285551662, dengan artian bahwa penanganan data outlier berhasil menurunkan nilai DBI dari sebelumnya 0.5293715978060519 yang menunjukkan hasil yang positif.

Dan cluster yang dihasilkan: Puskesmas di Cluster 1: B, C, D, E, F, G, H, I, J, L, O, P, R, S, T, U, V, W

Puskesmas di Cluster 2: K, N, Q, X

Puskesmas di Cluster 3: A, M,

SIMPULAN

Berdasarkan hasil analisis clustering dengan metode K-Means, diperoleh tiga kelompok Puskesmas di Kabupaten Asahan berdasarkan jumlah kasus stunting tahun 2022–2024. Setelah dilakukan deteksi dan penanganan outlier, nilai Davies-Bouldin Index menurun dari 0,529 menjadi 0,483, menunjukkan kualitas cluster yang lebih baik. Hasil akhir menunjukkan bahwa mayoritas Puskesmas masuk ke dalam Cluster 1 dengan kasus stunting relatif rendah hingga sedang, sedangkan Cluster 2 terdiri dari Puskesmas dengan tren kasus meningkat cukup signifikan, dan Cluster

2 berisi Puskesmas dengan jumlah kasus sangat tinggi.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang setinggi-tingginya kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia (Kemendikbudristek) atas dukungan pendanaan melalui skema Penelitian Dasar Pemula yang disetujui dalam Program Penelitian Nasional Tahun 2025.

DAFTAR PUSTAKA

1. De Onis M, Branca F. Childhood stunting: a global perspective. *Maternal & Child Nutrition*. 2016;12(S1):12-26.
2. Khan, L., & Hayat, A. (2024). Outlier detection in machine learning: A comprehensive review. *Journal of Machine Learning Research*, 25, 1-45.
3. Zhang, Z., & Zhao, X. (2024). Handling outliers in high-dimensional data: A new approach. *Data Mining and Knowledge Discovery*, 38(1), 12-34. doi:10.1007/s10618-023-00728-6.
4. Zhang C, Yang Q. A survey on multi-task learning. *IEEE Trans Knowl Data Eng*. 2023;35(1):1-20. doi:10.1109/TKDE.2022.3146592.
5. Gupta R, Sharma R. Enhancements in robust clustering methods for noisy data. *Data Mining and Knowledge Discovery*. 2024;38(1):45-67
6. Abeysinghe R, Abeywardena M, Ganewatta S. Applications of machine learning techniques in public health: A review. *J Health Inform Dev Ctries*. 2022;16(1):1-16.
7. Gonnelli S, Sassi F, Marini F. The Role of Machine Learning in Health Care: Applicability in Chronic Diseases. *Int J Environ Res Public Health*. 2021;18(3):1-15.
8. Alhussein M, Moderow T, Syed R. A comparative study of machine learning algorithms for assessing childhood malnutrition. *J Nutr Health Food Eng*. 2023;13(2):34-42.

-
9. Farhan M, Binsalman M, Beng H, Hussain M. Stunting prediction using advanced machine learning techniques: An analysis in rural settings. *BMC Public Health*. 2023;23:350.
 10. Yulianto, Setiawati R. Leveraging tree-based models for the prediction of stunting among children: Evidence from Indonesia. *Child Obes*. 2023;19(1):10-16.
 11. Lee J, Choi S. Exploring the intersection of machine learning and public health: A systematic review of current methods applied to stunting analysis. *BMC Health Serv Res*. 2023;23:558.
 12. Aji P, Suharyanto A, Evita M. Random Forests for identifying socio-economic determinants of child stunting in Indonesia. *Nutr Health*. 2023;29(1):67-76.
 13. Pramono M, Subekti A. The impact of economic factors on childhood stunting: An application of Gradient Boosting in rural Indonesia. *J Dev Studies*. 2023;59(4):650-670.
 14. Farhan M, Binsalman M, Beng H, Hussain M. Stunting prediction using advanced machine learning techniques: An analysis in rural settings. *BMC Public Health*. 2023;23:350.
 15. Abeysinghe R, Abeywardena M, Ganewatta S. Applications of machine learning techniques in public health: A review. *J Health Inform Dev Ctries*. 2022;16(1):1-16.
 16. Yulianto, Setiawati R. Leveraging tree-based models for the prediction of stunting among children: Evidence from Indonesia. *Child Obes*. 2023;19(1):10-16.
 17. Lee J, Choi S. Exploring the intersection of machine learning and public health: A systematic review of current methods applied to stunting analysis. *BMC Health Serv Res*. 2023;23:558.
 18. Sinuraya KA; Suwilo S; Lydia, MS. Accuracy Analysis on Images Retrieval System using Radial Basis Function Algorithm and Coefficient Correlation. 2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA). **2020**: 99-104.