ISSN: 2477-0078

Penerapan Metode K-Means untuk Klasterisasi Jumlah Balita Stunting Berdasarkan Data Puskesmas di Kabupaten Pamekasan

Tumiyah[1], Fathorrozi Ariyanto[2], Masdukil Makruf[3]
[1]Program Studi Sistem Informasi, Fakultas Teknik, Universitas Islam Madura
[2,3]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Madura
[1]mytumiyah@gmail.com, [2]fathorroziariyanto7@gmail.com, [3]masdukil.makruf@uim.ac.id

ABSTRAK

Stunting merupakan masalah pertumbuhan pada balita akibat kekurangan gizi dan infeksi berulang yang berdampak pada perkembangan fisik maupun kognitif anak. Di Kabupaten Pamekasan, isu ini masih memerlukan perhatian serius, namun belum ada kajian yang mengelompokkan wilayah berdasarkan tingkat keparahan kasus. Pemahaman pola jumlah stunting per puskesmas setiap tahun penting untuk menentukan prioritas intervensi secara efektif. Penelitian ini bertujuan mengelompokkan jumlah balita stunting di Kabupaten Pamekasan periode 2021–2024 menggunakan metode K-Means. Data diperoleh dari Dinas Kesehatan Kabupaten Pamekasan dan diolah melalui Google Colab. Tahapan analisis meliputi import data, normalisasi, serta penentuan jumlah klaster optimal dengan Silhouette Score, yang menghasilkan dua klaster dengan nilai 0,5308. Proses klasterisasi konvergen pada empat iterasi, menghasilkan 15 puskesmas pada klaster pertama dan 6 puskesmas pada klaster kedua. Evaluasi dengan Silhouette Coefficient (0,5182) menunjukkan kualitas klastering tergolong baik. Hasil ini menyediakan dasar data terstruktur untuk mendukung perencanaan intervensi stunting yang lebih efektif di Pamekasan.
Kata kunci: Balita Stunting, K-Means, Klastering, Puskesmas, Pamekasan

ABSTRACT

Stunting is a growth problem in toddlers due to malnutrition and repeated infections that impact children's physical and cognitive development. In Pamekasan Regency, this issue still requires serious attention, but no studies have grouped regions based on case severity. Understanding the pattern of stunting cases per community health center each year is important for determining effective intervention priorities. This study aims to cluster the number of stunted toddlers in Pamekasan Regency for the 2021–2024 period using the K-Means method. Data were obtained from the Pamekasan Regency Health Office and processed using Google Colab. The analysis stages included data import, normalization, and determining the optimal number of clusters using the Silhouette Score, which produced two clusters with a value of 0.5308. The clustering process converged in four iterations, resulting in 15 community health centers in the first cluster and 6 in the second cluster. Evaluation using the Silhouette Coefficient (0.5182) indicated that the clustering quality was relatively good. These results provide a structured data basis to support more effective stunting intervention planning in Pamekasan.
***Keywords**: Stunted Children, K-Means, Clustering, Health Centers, Pamekasan*

1. PENDAHULUAN

Stunting adalah gangguan pertumbuhan pada balita yang disebabkan oleh kekurangan gizi dalam jangka panjang serta infeksi berulang, yang berpengaruh pada perkembangan fisik maupun kemampuan kognitif anak (Rakhmalia Imeldawati, 2025)[1]. Kondisi ini tidak hanya menghambat peningkatan tinggi badan, tetapi juga memengaruhi perkembangan otak. Ketidakcukupan asupan nutrisi sejak usia dini dan frekuensi infeksi yang tinggi dapat menimbulkan keterlambatan dalam kemampuan berpikir, berbahasa, serta pengendalian motorik anak secara optimal (Lestari et al., 2024)[2].

Di Kabupaten Pamekasan, permasalahan stunting masih menjadi fokus utama dalam bidang kesehatan. Data dari Puskesmas

ISSN: 2477-0078

menunjukkan bahwa kasus ini terdapat di berbagai wilayah, seperti Kecamatan Galis, Palengaan, Proppo, serta kecamatan lain yang tersebar di seluruh kabupaten. Meskipun demikian, hingga kini belum ada kajian mendalam yang mengelompokkan wilayah berdasarkan tingkat keparahan kasus stunting. Analisis pola jumlah balita stunting setiap tahun di masing-masing Puskesmas diperlukan untuk menentukan wilayah prioritas intervensi secara lebih tepat dan efektif.

Penelitian ini bertujuan untuk mengelompokkan wilayah kerja Puskesmas di Kabupaten Pamekasan berdasarkan jumlah balita yang mengalami *stunting* setiap tahunnya selama periode 2021 hingga 2024 dengan menggunakan metode *Klastering K-Means*. Pengelompokan ini dilakukan untuk menemukan pola kesamaan antar Puskesmas terkait jumlah kasus *stunting* dari tahun ke tahun. Hasilnya diharapkan dapat membantu menentukan prioritas intervensi secara lebih efisien dan tepat sasaran.

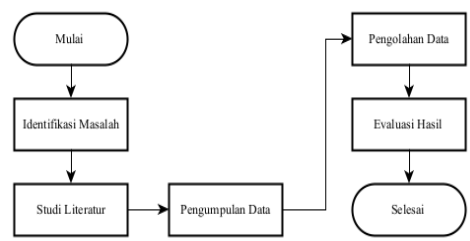
Sejumlah penelitian sebelumnya membuktikan bahwa metode K-Means efektif digunakan dalam analisis data kesehatan masyarakat, termasuk permasalahan gizi dan stunting. Nagari & Inayati (2020)[3] menerapkan K-Means untuk mengelompokkan status gizi balita di Ponkesdes Mayangrejo, Bojonegoro, dan berhasil membentuk klaster sesuai kondisi gizi yang berbeda. Selanjutnya, penelitian oleh Melisa & Syaiful Zuhri Harahap (2024)[4] memanfaatkan metode K-Means dan K-Medoids dalam pengelompokan risiko stunting balita di Puskesmas Sigambal. Di sisi lain, studi Wulandari dkk. (2021) menggunakan K-Means pada data keluhan kesehatan penduduk tingkat provinsi untuk mengidentifikasi wilayah dengan pola kesehatan serupa (Nurwanda et al., 2024)[5]. Namun, penerapan metode ini secara khusus untuk mengelompokkan jumlah balita stunting berdasarkan data tahunan riil di tingkat Puskesmas kabupaten, sebagaimana penelitian ini di Kabupaten Pamekasan, masih jarang dilakukan.

Penelitian ini menerapkan algoritma K-Means dengan penentuan pusat klaster (centroid) secara acak. Variabel yang dianalisis adalah jumlah balita stunting dari tiap Puskesmas pada tahun 2021–2024. Penentuan jumlah klaster dilakukan secara eksploratif, sedangkan kualitas hasil klastering dievaluasi menggunakan Silhouette Coefficient untuk melihat pemisahan antar klaster dan keseragaman data dalam satu klaster. Urgensi penelitian ini terletak pada penyediaan informasi berbasis data sebagai dasar pengambilan keputusan penurunan stunting di tingkat lokal. Melalui klasterisasi data riil Puskesmas, pemerintah daerah dapat lebih mudah mengenali wilayah dengan kasus stunting tinggi secara konsisten serta menyesuaikan strategi intervensi sesuai kondisi lapangan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode K-Means clustering untuk mengelompokkan jumlah balita stunting berdasarkan data tahunan Puskesmas di Kabupaten Pamekasan periode 2021–2024. Proses penelitian disusun melalui beberapa tahapan yang dirancang secara sistematis agar pelaksanaannya terarah, terencana, serta sesuai dengan tujuan yang ingin dicapai. Tahapan penelitian terlihat pada gambar 1 tersebut:



Gambar 1. Metodologi Penelitian

2.1.1 Identifikasi Masalah

Tahap ini berfokus pada perumusan masalah terkait penerapan metode K-Means clustering untuk mengelompokkan jumlah balita stunting berdasarkan data Puskesmas di Kabupaten Pamekasan, dengan mengidentifikasi permasalahan relevan yang akan dikaji lebih lanjut untuk menemukan solusi.

2.1.2 Studi Literatur

Tahap ini melibatkan kajian literatur terkait metode K-Means dan pemanfaatannya dalam analisis data kesehatan, termasuk stunting. K-Means terbukti efektif mengelompokkan data berdasarkan kemiripan untuk mengidentifikasi pola distribusi kasus dan wilayah prioritas.

ISSN: 2477-0078

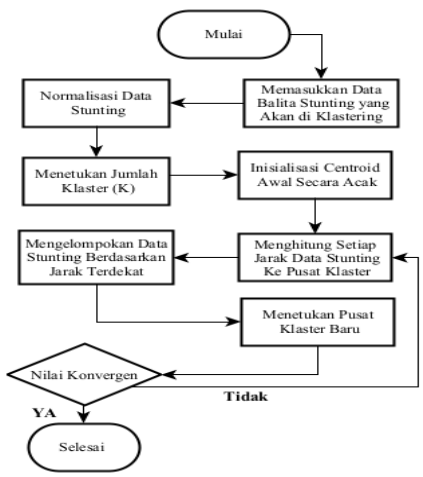
Studi literatur ini menjadi dasar teoritis, acuan pemilihan metode, serta bahan perbandingan dengan penelitian sebelumnya.

2.1.3 Pengumpulan Data

Data penelitian diperoleh dari Puskesmas di Kabupaten Pamekasan mengenai jumlah balita stunting tahun 2021–2024. Data sekunder ini bersumber dari Dinas Kesehatan (Dinkes) melalui observasi dan pencatatan. Pengelompokan dilakukan menggunakan metode Data Mining K-Means dengan fokus pada jumlah balita stunting. Setelah data terkumpul, analisis dilakukan untuk memberikan pemahaman yang lebih komprehensif terkait penelitian.

2.1.4 Pengolahan Data

Penelitian ini memanfaatkan Google Colab dan Microsoft Excel sebagai tools analisis data. Keduanya digunakan untuk penerapan metode K-Means karena mendukung integrasi pustaka data science serta memudahkan visualisasi dan pengolahan data. Adapun alur serta rumus K-Means yang digunakan dijelaskan sebagai berikut:



Gambar 2. Alur Pengolahan Data *K-Means*

Berdasarkan gambar di atas pengolahan data pada metode K Means ini akan melalui beberapa tahahapan diantaranya:

1) Memasukkan data balita stunting yang akan di klaster dan di olah menggunakan tools.
2) Normalisasi Data rumus yang di guanakan pada proses ini adalah sebagai berikut:

$$Z = \frac{X - \bar{X}}{s}$$

Keterangan:

z = nilai Z-Score (hasil normalisasi)
x = nilai asli
$\bar{x}$= rata-rata (mean) dari seluruh nilai dalam kolom
s = standar deviasi dari seluruh nilai dalam kolom

3) Menentukan jumlah klaster
4) Inisialisasi centroid secara acak yaitu memilih titik pusat awal (centroid) dari klaster secara acak dari data yang tersedia, sebelum proses pengelompokan dimulai.
5) Menghitung setiap jarak ke pusat klaster menggunakan rumus sebagai berikut:

$$d(x_i, C_k) = \sqrt{\sum_{j=1}^{n}(x_{ij} - C_{kj})^2}$$

Keterangan:
$d(xi,Ck)$ adalah jarak antara titik xi dan pusat klaster Ck
x_{ij} adalah nilai ke – j dari titik xi
C_{kj} adalah nilai atribut ke – j dari pusat klaster Ck
n adalah jumlah atribut

6) Mengelompokkan data berdasarkan jarak terdekat

$$\text{Cluster}(x_i) = \arg\min_j d(x_i, C_j)$$

Keterangan:
xi: Data ke-i yang akan dikelompokkan.
Cj: Centroid (pusat cluster) ke-j
d(xi ,Cj): Jarak antara data xi dengan centroid cluster ke-j, biasanya dihitung menggunakan jarak Euclidean.
argminj: Operator ini berarti memilih nilai j (index cluster) yang memberikan nilai minimum darid(xi ,Cj)

7) Menentukan pusat klaster baru atau Mengupdate posisi centroid

$$C_j = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_i$$

Keterangan:
Cj: Centroid baru untuk cluster ke-j.
Sj: Kumpulan data yang termasuk dalam cluster ke-j.
|Sj|: Jumlah data dalam cluster j.
xi: Data ke-i yang berada dalam cluster j.
∑: Penjumlahan semua data dalam cluster.

ISSN: 2477-0078

Proses K-Means dihentikan jika hasil iterasi sudah konvergen, namun jika berbeda dari iterasi sebelumnya, langkah perhitungan diulang hingga mencapai konvergensi.

2.1.5. Evaluasi Hasil

Tahap ini mengevaluasi hasil klastering jumlah balita stunting di Kabupaten Pamekasan tahun 2021–2024 menggunakan Silhouette Coefficient untuk menilai ketepatan jumlah klaster dan kualitas pengelompokan K-Means.

2.2 Data Mining

Data mining adalah proses analisis data berskala besar untuk menemukan pola, tren, dan informasi penting melalui teknik seperti klasifikasi, asosiasi, regresi, dan klasterisasi, yang mendukung pengambilan keputusan serta prediksi (M. Atif, 2022)[6]. Selain itu, data mining berperan penting dalam mengidentifikasi pola dan pengetahuan relevan dari data yang tersedia (Ananda Mustari et al., 2024)[7].

2.3 Klastering K-Means

Klastering adalah teknik pengelompokan data tanpa label berdasarkan tingkat kemiripan, di mana data dalam satu kelompok lebih serupa dibandingkan dengan kelompok lain (Nozomi, 2023)[8]. Salah satu algoritma yang banyak digunakan ialah K-Means, yang membagi data ke dalam klaster serupa berdasarkan jarak antar data (Joseph, 2023)[9]. K-Means bekerja dengan membagi data ke dalam *k* kelompok melalui perhitungan centroid, yaitu rata-rata data dalam tiap klaster (Sarifah Faujiah et al., 2024)[10]. Prosesnya dilakukan secara iteratif, dimulai dari pemilihan centroid acak, pengelompokan berdasarkan jarak terdekat, hingga pembaruan centroid sampai hasil stabil.

2.4 Metode Silhouette

Metode Silhouette digunakan untuk mengevaluasi kualitas hasil klastering dengan koefisien bernilai antara -1 hingga 1. Nilai mendekati 1 menunjukkan klaster baik, nilai mendekati 0 menandakan klaster kurang jelas, sedangkan nilai negatif menunjukkan kesalahan pengelompokan (R et al., 2022)[11]. Metode ini penting untuk menilai kekompakan dan pemisahan klaster, khususnya pada algoritma seperti K-Means yang membagi data berdasarkan jarak ke centroid (Fikri et al., 2023)[12].

Berikut kriteria penilaian dan rumus yang digunakan dalam evaluasi Silhouette:

Table 1. Kriteria *silhouette coeficient*

Rentang Nilai Silhouette Coefficient	Kriteria Penilaian
$0{,}7 \leq SC \leq 1{,}0$	Struktur Kuat
$0{,}5 < SC \leq 0{,}7$	Struktur Sedang
$0{,}25 < SC \leq 0{,}5$	Struktur Lemah
$SC \leq 0{,}25$	Tidak Ada Struktur

Source: [13]

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Keterangan:
a(i): jarak rata-rata ke klaster sendiri.
b(i): jarak rata-rata ke klaster terdekat lainnya.
Nilai s(i) dekat 1 → klaster bagus,
dekat 0 → di batas klaster,
negatif → salah klaster.

3. HASIL DAN PEMBAHASAN

1. Deskripsi Data Balita Stunting

Penelitian ini menggunakan data jumlah balita stunting di 21 Puskesmas Kabupaten Pamekasan periode 2021–2024 yang diperoleh dari Dinas Kesehatan. Jumlah kasus tercatat sebanyak 4.228 balita (2021), 2.885 (2022), 1.774 (2023), dan 1.880 (2024). Data yang digunakan berupa total kasus stunting per Puskesmas setiap tahun tanpa

ISSN: 2477-0078

mempertimbangkan variabel lain. Fokus penelitian adalah memetakan persebaran wilayah dengan beban stunting tinggi dan rendah untuk memberikan gambaran awal dalam penentuan prioritas intervensi percepatan penurunan stunting.

Gambar 3. Grafik Data Balita* Stunting *(Dinas Kesehatan)

Dalam klastering, data jumlah balita stunting per Puskesmas dikelompokkan berdasarkan jarak terdekat sesuai pola kemiripan. Puskesmas dengan jumlah kasus serupa masuk dalam satu klaster, sehingga hasilnya dapat membantu menentukan wilayah prioritas intervensi penurunan stunting.

2. Proses Klastering dengan Metode K-Means

Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan data jumlah balita stunting dari 21 Puskesmas di Kabupaten Pamekasan periode 2021–2024. Pengelompokan dilakukan berdasarkan kemiripan jumlah kasus dengan tujuan menemukan pola secara otomatis, menghasilkan dua kategori utama yaitu wilayah dengan kasus tinggi dan rendah. Jumlah klaster ditentukan menggunakan Silhouette Coefficient untuk menilai kualitas pengelompokan, sedangkan proses klastering mengikuti alur K-Means yang digambarkan pada flowchart.

1. Memasukkan Data yang Akan Diklaster

Pada tahap ini, data jumlah balita stunting per Puskesmas Kabupaten Pamekasan tahun 2021–2024 yang tercantum pada Tabel 2 dimasukkan ke dalam sistem untuk proses pengelompokan.

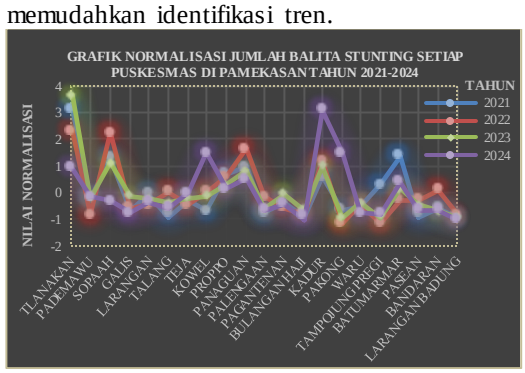
2. Normalisasi Data Balita *Stunting*

Pada tahap ini, data jumlah balita stunting yang telah dimasukkan ke sistem dinormalisasi agar proses klastering lebih konsisten, adil, dan memudahkan analisis tren.

```
[ ] from sklearn.preprocessing import StandardScaler
    scaler = StandardScaler()
    data_scaled = scaler.fit_transform(data_fitur)
    # 2. Normalisasi Data dengan StandardScaler

    display (data_scaled)

array([[ 3.12242766,  2.28725584,  3.6636443 ,  0.93497068],
       [-0.21436356, -0.83853721, -0.22228209, -0.15722333],
       [ 1.30235973,  2.23871868,  1.11312914, -0.30787078],
       [-0.63904608, -0.56672911, -0.12603624, -0.75981313],
       [-0.06875812, -0.4696548 , -0.23431283, -0.34553264],
       [-0.80285219,  0.02542423, -0.40274307, -0.58405777],
       [-0.33570142, -0.45994736, -0.19822063, -0.03168379],
       [-0.71184879,  0.01571679, -0.13806697,  1.48734467],
       [ 0.69567041,  0.58845527,  0.29503938,  0.14407157],
       [ 0.96261371,  1.64656534,  0.80033011,  0.5206902 ],
       [-0.76645083, -0.15901698, -0.63132698, -0.6593815 ],
       [-0.40243724, -0.59585141, -0.05385185, -0.42085637],
       [-0.85745423, -0.79970748, -0.65538844, -0.89790663],
       [ 0.56826566,  1.19031604,  1.00485256,  3.08169685],
       [-0.66938854, -1.14917502, -0.95615674,  1.49989862],
       [-0.52377511, -0.47936223, -0.40274307, -0.77236709],
       [ 0.28918857, -1.15888246, -0.85991088, -0.79747499],
       [ 1.38729623, -0.27550616,  0.1145784 ,  0.39515065],
       [-0.81498598, -0.29492102, -0.46289673, -0.64682754],
       [-0.58444404,  0.11279111, -0.70351137, -0.64682754],
       [-0.93632384, -0.85795207, -0.944126  , -1.03600012]])
```

Gambar 4. Hasil Normalisasi Data pada Google Colab

Tahap ini melakukan normalisasi data jumlah balita stunting yang telah dimasukkan ke sistem guna memastikan klastering berlangsung lebih konsisten, seimbang, dan memudahkan identifikasi tren.



Gambar 5. Grafik Hasil Normalisasi Data Balitas *Stunting*

3. Menentukan Jumlah Klaster

Jumlah klaster ditentukan melalui uji coba di Google Colab, dan hasilnya menunjukkan bahwa 2 klaster paling optimal untuk mengelompokkan puskesmas berdasarkan jumlah balita stunting.

```
Nilai Silhouette Coefficient SEBELUM K-means (dengan uji jumlah cluster 2-5):
  Jumlah Cluster = 2, Silhouette Coefficient = 0.5308
  Jumlah Cluster = 3, Silhouette Coefficient = 0.4798
  Jumlah Cluster = 4, Silhouette Coefficient = 0.4133
  Jumlah Cluster = 5, Silhouette Coefficient = 0.1845
```

Gambar 6. Nilai Optimal *Silhouette Coeficient*

4. Inisialisasi *Centroid* Awal

Centroid awal ditentukan secara acak, di mana peneliti memilih Puskesmas Galis pada baris ke-4 dan Puskesmas Teja pada baris ke-7 di Excel, serta baris ke-3 dan ke-6 pada Google Colab.

Tabel 2. Inisialisasi *Centroid* Awal

No	PUSKESMAS	2021	2022	2023	2024
4	GALIS	-0.639	-0.567	-0.126	-0.760
7	TEJA	-0.336	-0.460	-0.198	-0.032

```
manual_centroids = data_scaled[baris_centroid]
print(f"\nMenggunakan {jumlah_cluster} cluster dengan centroid awal dari baris: {baris_centroid}")
print("Centroid awal:")
print(manual_centroids)

Menggunakan 2 cluster dengan centroid awal dari baris: [3, 6]
Centroid awal:
[[-0.63904608 -0.56672911 -0.12603624 -0.75981313]
 [-0.33570142 -0.45994736 -0.19822063 -0.03168379]]
```

ISSN: 2477-0078

Gambar 7. Proses Inisialisasi *Centroid* Awal pada *Google Colab*

5. Menghitung Jarak dengan *Euclidean Distance*

Setelah menentukan centroid awal, jarak tiap data ke centroid dihitung secara manual di Microsoft Excel menggunakan fungsi **SQRT** sesuai rumus pada tahap ke-5.

Tabel 3. Menghitung Jarak *Eucludien Distance* (Iterasi 1)

PUSKESMAS	2021	2022	2023	2024	C1	C2
GALIS	-0.639	-0.567	-0.126	-0.760	0.000	0.799252
TALANG	-0.803	0.025	-0.403	-0.584	0.696	0.894853
PALENGAAN	-0.766	-0.159	-0.631	-0.659	0.669	0.926115
PAGANTENAN	-0.402	-0.596	-0.054	-0.421	0.421	0.441838
BULANGAN HAJI	-0.857	-0.800	-0.655	-0.898	0.633	1.160606
WARU	-0.524	-0.479	-0.403	-0.772	0.312	0.791322
TAMPOJUNG PREGI	0.289	-1.159	-0.860	-0.797	1.324	1.37959
PASEAN	-0.815	-0.295	-0.463	-0.647	0.481	0.839883
BANDARAN	-0.584	0.113	-0.704	-0.647	0.901	1.011742
LARANGAN BADUNG	-0.936	-0.858	-0.944	-1.036	0.959	1.443669
PADEMAWU	-0.214	-0.839	-0.222	-0.157	0.792	0.417603
LARANGAN	-0.069	-0.470	-0.234	-0.346	0.720	0.413711
TEJA	-0.336	-0.460	-0.198	-0.032	0.799	0
PAKONG	-0.669	-1.149	-0.956	1.500	2.477	1.872589
KOWEL	-0.712	0.016	-0.138	1.487	2.323	1.636707
TLANAKAN	3.122	2.287	3.664	0.935	6.287	5.94594
SOPAAH	1.302	2.239	1.113	-0.308	3.658	3.429571
PROPPO	0.696	0.588	0.295	0.144	2.027	1.561112
PANAGUAN	0.963	1.647	0.800	0.521	3.156	2.724929
KADUR	0.568	1.190	1.005	3.082	4.537	3.831586
BATUMARMAR	1.387	-0.276	0.115	0.395	2.363	1.811842

6. Mengelompokkan Data *Stunting* Berdasarkan Jarak Terdekat

Pada tahap ini, data dikelompokkan ke dalam 2 klaster berdasarkan jarak terdekat dengan centroid, sebagaimana ditampilkan pada tabel berikut.

Tabel 4. Pengelompokan Data Berdasarkan Jarak Terdekat (Iterasi 1)

PUSKESMAS	2021	…	C1	..	JARAK TERDEKAT	KLASTER
GALIS	-0.639	…	0.000	..	0	1
TALANG	-0.803	…	0.696	..	0.696372486	1
PALENGAAN	-0.766	…	0.669	..	0.669228213	1
PAGANTENAN	-0.402	…	0.421	..	0.420635378	1
BULANGAN HAJI	-0.857	…	0.633	..	0.633454502	1
WARU	-0.524	…	0.312	..	0.312481427	1
TAMPOJUNG PREGI	0.289	…	1.324	..	1.323727899	1
PASEAN	-0.815	…	0.481	..	0.480702841	1
BANDARAN	-0.584	…	0.901	..	0.900540055	1
LARANGAN BADUNG	-0.936	…	0.959	..	0.958506665	1
PADEMAWU	-0.214	…	0.792	..	0.417603011	2
LARANGAN	-0.069	…	0.720	..	0.41371709	2
TEJA	-0.336	…	0.799	..	0	2
PAKONG	-0.669	…	2.477	..	1.872588675	2
KOWEL	-0.712	…	2.323	..	1.636706801	2
TLANAKAN	3.122	…	6.287	..	5.94594011	2
SOPAAH	1.302	…	3.658	..	3.429571434	2
PROPPO	0.696	…	2.027	..	1.561112207	2
PANAGUAN	0.963	…	3.156	..	2.724929093	2
KADUR	0.568	…	4.537	..	3.831585818	2
BATUMARMAR	1.387	…	2.363	..	1.811841695	2

Tabel 6 menyajikan data yang telah diurutkan sesuai hasil klaster, sehingga tiap puskesmas ditempatkan pada klaster masing-masing.

7. Menentukan Pusat Klaster Baru

Setelah klaster terbentuk, centroid baru dihitung dari rata-rata data tiap kelompok, lalu digunakan kembali untuk perhitungan jarak dan pengelompokan.

Tabel 5. Pusat Klaster baru pada iterasi terakhir

Centroid	2021	2022	2023	2024
C1	-0.5359	-0.5117	-0.4661	-0.3179
C2	1.3398	1.2793	1.1653	0.7948

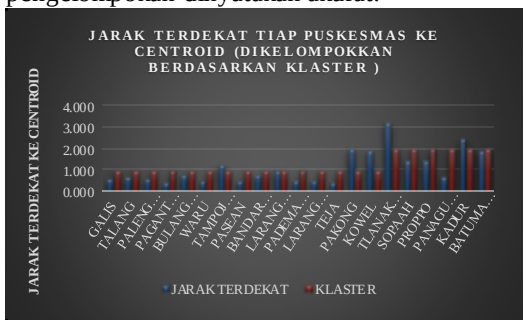
Pada penelitian ini melakukan 4 kali iterasi hingga centroid mencapai kondisi konvergen atau stabil, baik pada perhitungan manual maupun di Google Colab.

```
[ ] print(f"\nJumlah iterasi yang dilakukan: {kmeans.n_iter_}")
    print("\nPosisi centroid akhir untuk masing-masing cluster:")
    for i, center in enumerate(kmeans.cluster_centers_):
        print(f"  Cluster {i}: {center}")

[4]
Jumlah iterasi yang dilakukan: 4

Posisi centroid akhir untuk masing-masing cluster:
  Cluster 0: [-0.53590889 -0.51172033 -0.46610493 -0.31791394]
  Cluster 1: [1.33977223 1.27930084 1.16526231 0.79478486]
```

Gambar 8. Proses iterasi dan posisi *Centroid* akhir masing-masing klaster

8. Hasil *Klastering*

Tahap akhir penelitian menghasilkan klastering setelah 4 iterasi. Perhitungan manual di Excel dan pengolahan di Google Colab memberikan hasil yang sama, sehingga proses pengelompokan dinyatakan akurat.



Gambar 9. Grafik Hasil *Klastering K-Means*

Hasil klasterisasi menghasilkan dua kelompok puskesmas. Klaster 1 (centroid [-0,5359, -0,5117, -0,4661, -0,3179]) beranggotakan 15 puskesmas dengan karakteristik angka stunting rendah dan stabil. Klaster 2 (centroid [1,3397, 1,2793, 1,1652, 0,7947]) mencakup 6

ISSN: 2477-0078

puskesmas dengan kasus stunting lebih tinggi atau fluktuatif, sehingga memerlukan intervensi lebih intensif.

Pengelompokan dilakukan menggunakan jarak Euclidean, di mana data ditempatkan pada klaster dengan jarak terdekat (Lubis & Hasibuan, 2024)[14]. Secara deskriptif, Klaster 1 menggambarkan wilayah dengan akses layanan kesehatan lebih baik, sedangkan Klaster 2 mencerminkan wilayah dengan kondisi sedang dan prevalensi stunting lebih signifikan (Ramadhani et al., 2025)[15].

3. Evaluasi *Silhouette Coefficient* dan Evaluasi Kualitas *Klastering*

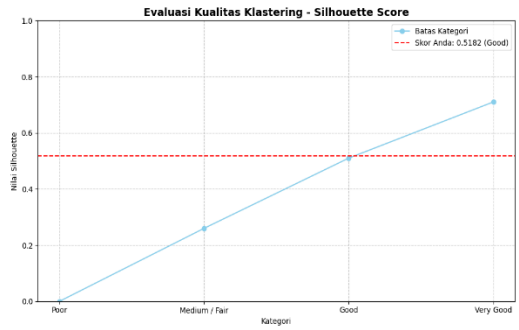
Setelah klasterisasi selesai, kualitas hasil dievaluasi menggunakan **Silhouette Coefficient**, yaitu ukuran kesesuaian data dengan klasternya dibandingkan klaster lain, dengan rentang nilai -1 hingga 1. Penelitian ini memperoleh nilai **0,5182**, yang menunjukkan pemisahan antar kelompok cukup jelas serta kekompakan dalam tiap klaster. Nilai tersebut termasuk kategori *Good*, sehingga hasil klasterisasi dinilai memadai untuk analisis lebih lanjut.

```
from sklearn.metrics import silhouette_score
score = silhouette_score(data_scaled, cluster_labels)
print(f"\nSilhouette Coefficient: {score:.4f}")

if score >= 0.71:
    kategori = "Very Good"
elif score >= 0.51:
    kategori = "Good"
elif score >= 0.26:
    kategori = "Medium / Fair"
elif score >= 0.0:
    kategori = "Poor"
else:
    kategori = "Salah Cluster (nilai negatif)"

print(f"Evaluasi Kualitas Klastering: {kategori}")
```

```
Silhouette Coefficient: 0.5182
Evaluasi Kualitas Klastering: Good
```

Gambar 10. Hasil *Silhouette Coefficient* dan Hasil Evaluasi Kualitas Klaster

Sebagai bagian dari validasi, evaluasi **Silhouette Coefficient** juga dihitung manual di Microsoft Excel menggunakan rumus yang sesuai. Hasilnya konsisten dengan perhitungan di Google Colab, sehingga memperkuat validitas klasterisasi dan menegaskan bahwa penggunaan dua klaster dalam pengelompokan data balita stunting di Kabupaten Pamekasan sudah tepat.



Gambar 11. Grafik Hasil Evaluasi *Klastering K-Means* Mengguakan *Silhouette Coefficient*

Hasil evaluasi dengan **Silhouette Coefficient** sebesar 0,5182 menunjukkan klastering sudah baik, dengan data yang homogen dalam klaster dan jelas terpisah antar klaster. Struktur ini dinilai efektif serta dapat dijadikan dasar strategi intervensi stunting di Kabupaten Pamekasan.

4. KESIMPULAN

Penelitian ini mengelompokkan data balita stunting di Kabupaten Pamekasan menggunakan algoritma K-Means. Setelah normalisasi data dan uji Silhouette Score, diperoleh dua klaster optimal dengan konvergensi pada 4 iterasi: Klaster 1 berisi 15 puskesmas dan Klaster 2 berisi 6 puskesmas. Evaluasi dengan Silhouette Coefficient sebesar 0,5182 menunjukkan kualitas klastering baik, sehingga hasil ini dapat dijadikan dasar analisis dan perencanaan intervensi stunting di Pamekasan.

UCAPAN TERIMA KASIH

Peneliti menyampaikan terima kasih kepada Dinas Kesehatan Kabupaten Pamekasan atas penyediaan data, kepada orang tua dan keluarga atas doa serta dukungan, kepada dosen pembimbing atas arahan, serta kepada rekan-rekan dan semua pihak yang telah membantu. Semoga penelitian ini bermanfaat dan menjadi referensi dalam penanganan stunting ke depan.

DAFTAR PUSTAKA

[1] Rakhmalia Imeldawati, “Dampak

ISSN: 2477-0078

Terjadinya Stunting terhadap Perkembangan Kognitif Anak : Literature Review," *J. Med. Nusant.*, vol. 3, no. 1, pp. 101–107, 2025, doi: 10.59680/medika.v3i1.1632.

[2] E. Lestari, A. Siregar, A. K. Hidayat, and A. A. Yusuf, "Stunting and its association with education and cognitive outcomes in adulthood: A longitudinal study in Indonesia," *PLoS One*, vol. 19, no. 5, pp. 1–18, 2024, doi: 10.1371/journal.pone.0295380.

[3] S. S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method To Determine Nutritional Status," *J. Biometrika dan Kependud.*, vol. 9, no. 1, pp. 62–68, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.

[4] dan M. Melisa, Syaiful Zuhri Harahap, "Implementasi Data Mining untuk Klustering Stunting Gizi pada Balita di Puskesmas Sigambal Menggunakan Metode K-Medoids dan K-Means," *Sport. Cult.*, vol. 15, no. 1, pp. 72–86, 2024, doi: 10.25130/sc.24.1.6.

[5] M. D. Nurwanda, A. N. Liviasari, and D. D. Pambudi, "Penerapan K-Means untuk Meningkatkan Strategi Pemasaran Melalui Segmentasi Rumah Tangga yang Efektif," *J. Penelit. Ilmu …*, 2024, [Online]. Available: https://jpilkom.org/index.php/journal/article/view/57

[6] M. Atif, "Data Mining Data mining," *Min. Massive Datasets*, vol. 2, no. January 2013, pp. 5–20, 2022, [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

[7] K. Ananda Mustari, P. Assiroj, B. Hartati, and F. Samuel, "Implementasi Data Mining Pada Instansi Pemerintahan (Systematic Literature Review)," *J. Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3137–3142, 2024.

[8] I. Nozomi, "Penerapan Data Mining Untuk Peringatan Dini Banjir Menggunakan Metode Klastering K-Means (Studi Kasus Kota Padang)," *J. Sains Inform. Terap.*, vol. 2, no. 2, pp. 39–44, 2023, doi: 10.62357/jsit.v2i2.165.

[9] S. Joseph, "Development of a clustering algorithm for universal color image segmentation," no. July, 2023, [Online]. Available: https://openscholar.dut.ac.za/handle/10321/4787

[10] R. Sarifah Faujiah, C. Naya, and A. Susilo, "Analisis Klustering Peningkatan Jumlah Produksi Baterai Lithium-Ion dengan Algoritma Metode K-Means," *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 6, no. 5, pp. 1906–1914, 2024, doi: 10.38035/rrj.v6i5.1021.

[11] N. N. F. R, D. S. Anggraeni, and U. Enri, "Pengelompokkan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient," *Tematik*, vol. 9, no. 1, pp. 29–35, 2022, doi: 10.38204/tematik.v9i1.901.

[12] A. Fikri, B. F. Hutabarat, and U. Khaira, "Komparasi Antara Metode K-Means Clustering Dan Complete Linkage Dalam Pengelompokan Penyaluran Pinjaman Oleh Financial Technology," *J. Ilm. Media Sisfo*, vol. 17, no. 2, pp. 228–239, 2023, doi: 10.33998/mediasisfo.2023.17.2.1373.

[13] N. Rohman and A. Wibowo, "Clustering of Popular Spotify Songs in 2023 Using K-Means Method and Silhouette Coefficient," *J. Pilar Nusa Mandiri*, vol. 20, no. 1, pp. 18–24, 2024, doi: 10.33480/pilar.v20i1.4937.

[14] M. T. H. Lubis and M. S. Hasibuan, "Measurement of Centroid Distance in Determining Stunting Clusters," *J. La Multiapp*, vol. 5, no. 3, pp. 270–285, 2024, doi: 10.37899/journallamultiapp.v5i3.1479.

[15] F. Ramadhani, D. Septiana, S. N. Amalia, P. Maulidina, and A. Satria, "Data Science : Journal Of Computing And Applied Informatics Spatial

ISSN: 2477-0078

Clustering Analysis of Stunting in North Sumatra Based on Environmental Factors Using K-Means Algorithm," vol. 9, no. 2, pp. 18–25, 2025, doi: 10.32734/jocai.v9.i2-17179.