

Kombinasi Principal Component Analysis dengan Algoritma K-Means untuk Klasterisasi Data Stunting

Gladys Fouriza Ibanez*, Giri Wahyu Wiriasto, Rosmaliati

Fakultas Teknik, Program Studi Teknik Elektro, Universitas Mataram, Mataram, Indonesia
Email: ^{1*}gladysfouriza@gmail.com, ²giriwahyuwiriasto@gmail.com, ³rosmaliati.muchtar@gmail.com
Email Penulis Korespondensi: gladysfouriza@gmail.com

Abstrak—Kabupaten Dompu di Nusa Tenggara Barat merupakan kabupaten dengan prevalensi stunting balita tertinggi ke-4 di provinsi NTB pada tahun 2022, yakni 34.5%. Penelitian ini menggunakan data dari Dinas Kesehatan Kabupaten Dompu tahun 2023 yaitu data dari 81 Desa yang mencakup enam variabel faktor yang mempengaruhi tingkat prevalensi stunting. Tujuan dari penelitian ini menentukan dan mengetahui karakteristik suatu kawasan berdasarkan faktor stunting di Kabupaten Dompu dengan metode K-Means Clustering yang dikombinasikan dengan metode *Principal Component Analysis* (PCA). Metode K-Means menghasilkan cluster berupa pembagian kawasan yang memiliki karakteristik yang berbeda, karakteristik diperoleh berdasarkan hasil reduksi data dari PCA yaitu berupa komponen utama. Optimalisasi jumlah Cluster dengan metode *Elbow* menunjukkan 3 Cluster yaitu berupa kawasan dengan Zona Tipe 1 yang terdiri dari 42 Desa, Zona Tipe 2 yang terdiri dari 12 Desa dan Zona Tipe 3 yang terdiri dari 27 Desa, kemudian dilakukannya tahapan evaluasi menggunakan metode *Silhouette* yang menghasilkan 2 Cluster yaitu berupa kawasan dengan Zona Tipe 1 yang terdiri dari 54 Desa, dan Zona Tipe 2 yang terdiri dari 22 Desa dengan nilai *Silhouette Score* 0.53, menunjukkan struktur model Cluster yang cukup baik. PCA menghasilkan dua komponen utama dengan nilai eigen tertinggi, masing-masing menjelaskan variansi 58.7% dan 14.28%, dengan kumulatif variansi mencapai 72.9%. Hal ini menunjukkan bahwa dua komponen utama ini dapat menunjukkan faktor yang mempengaruhi tingkat prevalensi stunting di Kabupaten Dompu.

Kata Kunci: Stunting; K-Means; Principal Component Analysis; Metode Elbow; Metode *Silhouette*

Abstract—Dompu Regency in West Nusa Tenggara ranks as the fourth highest regency in terms of stunting prevalence among toddlers in the NTB province in 2022, with a rate of 34.5%. This study utilizes data from the Dompu Regency Health Office in 2023, covering 81 villages and six variables that influence the prevalence of stunting. The objective of this research is to determine and understand the characteristics of areas based on stunting factors in Dompu Regency using the K-Means Clustering method combined with Principal Component Analysis (PCA). The K-Means method produces clusters that represent areas with different characteristics, derived from the data reduction results of PCA, which form the principal components. The optimization of the number of clusters using the Elbow method indicates 3 clusters, consisting of Zone Type 1 with 42 villages, Zone Type 2 with 12 villages, and Zone Type 3 with 27 villages. Subsequently, an evaluation phase using the *Silhouette* method resulted in 2 clusters: Zone Type 1 with 54 villages and Zone Type 2 with 22 villages, with a *Silhouette Score* of 0.53, indicating a fairly good cluster structure. PCA produced two principal components with the highest eigenvalues, each explaining 58.7% and 14.28% of the variance, with a cumulative variance of 72.9%. This demonstrates that these two principal components can effectively represent the factors influencing stunting prevalence in Dompu Regency.

Keywords: Stunting; K-Means; Principal Component Analysis; Elbow Method; *Silhouette* Method

1. PENDAHULUAN

Stunting adalah masalah kurang gizi kronis akibat kurangnya asupan gizi dalam waktu yang cukup lama sehingga mengakibatkan gangguan pertumbuhan pada anak. Seorang anak dianggap mengalami stunting jika tinggi badan mereka lebih rendah atau pendek (kerdil) dari standar usianya [1]. Menurut Kementerian Kesehatan 2018, Stunting dapat terjadi akibat kekurangan gizi terutama pada saat 1000 Hari Pertama Kehidupan (HPK). Balita yang menderita stunting akan lebih mudah terkena penyakit, juga berpotensi menghambat pertumbuhan fisik serta perkembangan kognitif yang akan berpengaruh terhadap tingkat kecerdasan dan produktivitas anak di masa mendatang.

Hasil Survei Status Gizi Indonesia (SSGI) 2022 menunjukkan bahwa angka kejadian stunting secara nasional sebesar 21,6 %. Prevalensi balita stunting mengacu pada persentase jumlah balita di suatu populasi yang mengalami stunting dalam pertumbuhan fisiknya. Berdasarkan laporan SSGI provinsi Nusa Tenggara Barat 2022, Kabupaten Dompu merupakan Kabupaten dengan tingkat prevalensi balita stunting tertinggi ke-4 di provinsi NTB dengan prevalensi sebesar 34.5% setelah Kabupaten Lombok Tengah sebesar 37.0%, Kabupaten Lombok Utara 35.9%, dan Kabupaten Lombok Timur 35.6%.

Menghadapi tingginya prevalensi stunting di Kabupaten Dompu penelitian ini bertujuan untuk menentukan dan mengetahui karakteristik suatu kawasan berdasarkan faktor stunting di Kabupaten Dompu dengan metode K-Means Clustering yang dikombinasikan dengan metode *Principal Component Analysis* (PCA). Pendekatan ini dilakukan dengan menganalisis faktor penyebab tingginya prevalensi balita stunting yang menjadi karakteristik dari hasil pengelompokan kawasan di Kabupaten Dompu.

Metode *Principal Component Analysis* atau PCA merupakan suatu teknik multivariat yang bertujuan untuk mereduksi faktor atau variabel dalam jumlah besar menjadi beberapa faktor yang lebih sedikit. Permasalahan yang sering muncul dalam proses reduksi faktor atau variabel ini yakni bagaimana memperkecil jumlah variabel namun tetap mempertahankan informasi atau karakter penting yang terkandung pada data [2].

K-means adalah salah satu algoritma *Data Mining* yang populer digunakan untuk menganalisis data dengan mengelompokkan data ke dalam *Cluster* yang memiliki kesamaan berdasarkan atribut-atributnya serta merupakan algoritma yang paling mudah dipahami sehingga metode ini sangat cocok digunakan dalam mengelompokkan suatu

kawasan berdasarkan kedekatan datanya. Konsep dasar dari algoritma K-means adalah mempartisi data ke dalam *Cluster*, di mana *Cluster* adalah jumlah kelompok yang dihasilkan oleh algoritma K-means [3].

Berdasarkan penelitian yang dilakukan oleh Sintha dan Aji (2023) menggunakan metode analisis PCA yaitu untuk mengetahui faktor dominan berdasarkan 17 faktor yang menentukan kelulusan studi mahasiswa [4]. Penelitian yang dilakukan oleh Kautsar dan Arie (2021) membuktikan metode PCA dapat memaksimalkan performa model *clustering* dalam menentukan karakterisasi kemiskinan di Indonesia pada tahun 2021, dimana model *clustering* menggunakan PCA memiliki nilai *average silhouette* yang lebih tinggi daripada model tanpa PCA. Hal ini menunjukkan bahwa masalah *curse of dimensionality* pada *clustering* dapat diatasi dengan menggunakan PCA [5].

Sementara itu, penelitian yang dilakukan oleh Sugiyarto, Khang, Choo, dan Ferna (2023) yaitu melakukan *Cluster* dengan algoritma K-Means dan reduksi dimensi data dengan PCA yang bertujuan untuk meningkatkan akurasi model prediksi. PCA mereduksi dimensi data agar data yang memiliki banyak variabel dapat lebih mudah dianalisis [6].

Penggabungan metode K-Means *Clustering* dan PCA juga dilakukan dalam penelitian Istina dan Dina (2024) untuk mengurangi masalah variabel yang kompleks berdasarkan tujuh variabel indikator kesejahteraan masyarakat Provinsi Jawa Barat. Hasil dari penelitian tersebut yaitu terdapat dua komponen utama berdasarkan nilai *eigenvalues*,serta membentuk 3 *cluster* terbaik [7].

Penelitian yang dilakukan oleh Riva, Rini, dan Yuyun (2021) membahas penerapan K-Means dan K-Medoids *Clustering* dalam perencanaan kebutuhan obat-obatan di puskesmas. Hasil penelitian ini menunjukkan bahwa K-Means memberikan *Silhouette Coefficient* yang lebih tinggi, menandakan bahwa hasil *Clustering* dengan menggunakan algoritma K-Means lebih baik dibandingkan K-Medoids. Dari hasil pengelompokan obat-obatan, puskesmas dapat melakukan pemerataan stok obat untuk menghindari kelebihan atau kekurangan stok [8].

Berdasarkan penelitian yang dilakukan oleh Agus, Rifqi, dan Achmad (2024) yaitu menggunakan algoritma K-Means untuk mengonfirmasi pentingnya memahami kebutuhan pelanggan, dengan mengelompokkan data secara tepat, dan mengambil tindakan efektif untuk menjaga kepuasan pelanggan. Jumlah *Cluster* yang ditetapkan berdasarkan nilai dari *Davies Bouldin* membentuk 2 *Cluster* baru [9].

Penelitian yang dilakukan oleh Desi, Budi, Rini, dan Apri (2024) menggunakan metode siku (*Elbow method*) untuk membantu menentukan jumlah *Cluster* terbaik yang akan dibuat oleh pemodelan K-Means dalam proses segmentasi *customer*. Berdasarkan metode *Elbow* menghasilkan jumlah *Cluster* yang optimal adalah 3 *Cluster*. Titik yang menunjukkan *Cluster* optimal dalam metode ini akan membentuk siku dalam sebuah grafik [10].

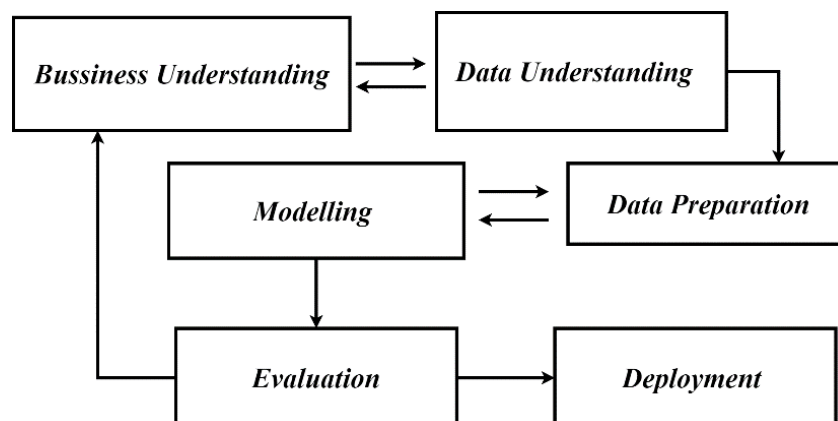
Menentukan nilai K optimal dan mengevaluasi hasil *clustering* dapat dilakukan dengan menggunakan metode *silhouette*. Penelitian yang dilakukan oleh Suraya, Muhamad, dan Uning (2023) yaitu evaluasi *Silhouette* dengan pengelompokan terbaik jumlah *cluster* optimal $k = 2$ dan memiliki *Silhouette Score* 0.656 [11].

Penelitian yang dilakukan oleh Erlisa dan Dede (2024) juga melakukan evaluasi hasil *Clustering* berdasarkan nilai *Silhouette score*. Hasil evaluasi digunakan untuk menilai kualitas *Cluster* metode K-Means. *Silhouette score* yang dihasilkan bernilai 0.617 menunjukkan bahwa kualitas *Cluster* dapat dikatakan baik [12].

Maka berdasarkan uraian tersebut, melalui penelitian ini penulis akan mengkaji mengenai *Clustering* Kabupaten Dompu di provinsi Nusa Tenggara Barat dengan mengelompokkan kawasan menjadi zona yang memiliki karakteristik berdasarkan faktor penyebab tingginya tingkat prevalensi stunting pada balita menggunakan algoritma K-Means yang dikombinasikan dengan PCA. Hasil pengelompokan zona kemudian diidentifikasi sebagai kawasan yang menjadi referensi bagi pemerintah dalam mengambil kebijakan yang sesuai terkait penurunan prevalensi stunting di Kabupaten Dompu.

2. METODOLOGI PENELITIAN

Metode yang digunakan pada penelitian ini yaitu mengikuti tahapan yang terdapat dalam proses Data Mining dengan menggunakan tahapan CRISP-DM (Cross-Industry Standard Process for Data Mining) yang dibangun dari 6 tahapan yang direpresentasikan melalui rancangan penelitian pada Gambar 1



Gambar 1. Tahapan CRISP-DM

2.1 Business Understanding

Pada tahapan ini memahami dan menganalisis tingkat stunting pada balita di Kabupaten Dompu. Analisis dilakukan berdasarkan data yang didapatkan dari Dinas Kesehatan Kabupaten Dompu pada tahun 2023 mengenai faktor-faktor yang mempengaruhi tingkat prevalensi stunting yang mencakup aspek-aspek seperti kesehatan ibu dan anak, gizi dan imunisasi, serta kesehatan sanitasi. Selanjutnya yaitu memilih metode *Clustering* untuk analisis data, menggunakan algoritma K-Means dengan pendekatan *Principal Component Analysis* (PCA) serta menetapkan Google Colaboratory sebagai lingkungan kerja.

2.2 Data Understanding

Variabel yang digunakan dalam penelitian ini yaitu faktor yang mempengaruhi tingkat prevalensi stunting, antara lain :

a. Hubungan Riwayat Konsumsi TTD Saat Kehamilan Ibu dengan Stunting

Langkah awal untuk mengantisipasi balita *stunting* dapat dilakukan dengan memperhatikan gizi ibu pada saat kehamilan. Kurangnya zat besi (Fe) selama kehamilan dapat meningkatkan risiko terjadinya anemia. Anemia tersebut dapat menyebabkan risiko kelahiran bayi prematur dan bayi dengan berat badan lahir rendah (BBLR). Ibu yang mendapatkan gizi dengan secara patuh mengonsumsi Tabelt Fe memiliki risiko lebih kecil untuk terjadinya *stunting* pada anak jika dibandingkan dengan ibu yang tidak patuh mengonsumsi Tabelt Fe [13].

b. Hubungan Inisiasi Menyusui Dini (IMD) dengan Stunting

Inisiasi Menyusui Dini (IMD) mempengaruhi kejadian stunting karena dengan IMD bayi akan mendapatkan ASI pertama kali yang mengandung kolostrum yang tinggi kaya akan antibodi dan zat penting untuk pertumbuhan usus dan ketahanan terhadap infeksi yang sangat dibutuhkan bayi demi kelangsungan hidupnya. Bayi yang IMD lebih tahan terhadap infeksi sehingga kemungkinan untuk sakit sangat kecil sehingga jumlah asupan tidak terganggu karena asupan merupakan faktor langsung yang menentukan status gizi balita [14].

c. Hubungan ASI Eksklusif dengan Stunting

Air Susu Ibu (ASI) eksklusif dapat mempengaruhi kejadian stunting karena jika bayi yang belum cukup umur 6 bulan sudah diberi makanan selain ASI akan menyebabkan usus bayi tidak mampu mencerna makanan dan bayi akan mudah terkena penyakit karena kurangnya asupan. Sehingga balita yang sering menderita penyakit infeksi akan menyebabkan pertumbuhannya terhambat dan tidak dapat mencapai pertumbuhan yang optimal. Manfaat dari ASI salah satunya adalah mendukung pertumbuhan bayi terutama tinggi badan karena kalsium ASI lebih efisien diserap dibanding susu pengganti ASI atau susu formula. Sehingga bayi yang diberikan ASI eksklusif cenderung memiliki tinggi badan yang lebih tinggi dan sesuai kupa pertumbuhan dibandingkan dengan bayi yang diberikan susu formula [15].

d. Hubungan Pemberian Vitamin A pada Bayi dengan Stunting

Vitamin A adalah zat gizi mikro yang sangat penting. Kekurangan vitamin A mempengaruhi sintesis protein, sehingga memengaruhi pertumbuhan sel dan diferensiasi. Di Indonesia, pemberian suplemen vitamin A telah dilaporkan dapat menurunkan angka kematian dan angka kejadian penyakit. Pemenuhan konsumsi vitamin A sangat penting untuk mendukung tumbuh kembang balita, maka disarankan untuk ibu yang mempunyai balita untuk memberikan vitamin A yang bisa di dapat di fasilitas kesehatan [16].

e. Hubungan Bayi Menerima Imunisasi Dasar Lengkap dengan Stunting.

Imunisasi adalah suatu cara untuk memberikan kekebalan tubuh terhadap seseorang secara aktif terhadap penyakit menular. Anak yang tidak mendapatkan imunisasi secara lengkap lebih rentan terkena penyakit, khususnya penyakit diare dan ISPA sehingga anak kesulitan atau tidak nafsu makan yang berdampak tidak terpenuhinya gizi pada balita yang pada akhirnya menyebabkan anak memiliki tubuh yang lebih pendek dari yang seharusnya [17].

f. Hubungan Akses Terhadap Fasilitas Sanitasi yang Layak (Jamban Sehat) dengan Stunting.

Jamban yang sehat adalah yang memenuhi persyaratan kesehatan guna mencegah tersebarnya sumber penyakit. Anak yang mana memiliki sanitasi lingkungan kurang baik akan memiliki risiko mengalami stunting dibandingkan anak yang sanitasi lingkungan cukup dan baik. Kepemilikan jamban yang tidak memenuhi syarat atau tidak layak memiliki risiko atau berpeluang terjadinya penyakit infeksi yang dapat menyebabkan terganggunya serapan nutrisi pada proses pencernaan sehingga berdampak pada penurunan berat badan bayi. Anak yang menggunakan jamban tidak sehat akan nerisiko mengalami stunting [18].

2.3 Data Preparation

Proses persiapan data dimulai dengan memilih data yang relevan untuk di analisis. Langkah berikutnya adalah mengidentifikasi nilai yang hilang atau missing value, karena nilai yang hilang dapat mempengaruhi hasil analisis, kemudian melakukan pengecekan data yang terduplikat, selanjutnya data di transformasi menggunakan teknik penskala untuk persiapan analisis.

2.4 Modelling

Pada tahap ini, sebelum melakukan pemodelan perlu diketahui terlebih dahulu jumlah *cluster* yang optimal. Penentuan jumlah *cluster* ini dapat menggunakan metode *Elbow*. Metode *Elbow* merupakan salah satu metode yang dapat digunakan

untuk menentukan jumlah *Cluster* terbaik, yaitu dengan cara melihat presentase setiap *Cluster* yang akan membentuk siku pada suatu titik tertentu. Metode *Elbow* biasa disajikan dalam bentuk grafik untuk mengetahui lebih jelas siku yang terbentuk. Tujuan dari metode *Elbow* adalah untuk memilih nilai *k* yang kecil dan masih memiliki nilai *withinss* yang rendah [19]. Setelah jumlah *cluster* optimal sudah diketahui maka selanjutnya yaitu tahapan pemodelan *clustering* menggunakan algoritma K-Means,

2.5 Evaluation

Pada tahap ini dilakukannya evaluasi kualitas model yang dibangun menggunakan metrik *silhouette coefficient*. Metode *Silhouette* digunakan untuk memilih jumlah *Cluster* optimal dengan menggunakan data skala rasio. Ketika diterapkan, algoritma *silhouette* akan mengukur jarak rata-rata dari suatu objek terhadap seluruh objek yang terdapat pada *Cluster* yang sama dengan objek di *Cluster* lainnya. Jika nilai indeks *Silhouette* mendekati 1 berarti proses *Clustering* kuat/baik, dan jika mendekati 0 berarti prosedur *Clustering* tidak cukup baik [20].

Tabel 1 Interpretasi *Silhouette Coefficient*

<i>Silhouette Coefficient</i>	Interpretasi
0.71-1.00	<i>Cluster</i> Kuat
0.51-0.70	<i>Cluster</i> Layak/Sesuai
0.26-0.50	<i>Cluster</i> Lemah
<0.25	Tidak dapat disebut <i>Cluster</i>

2.6 Deployment

Menghasilkan laporan akhir yang menjelaskan proses dan hasil penelitian yang mencakup semua tahapan yang telah dilakukan, mulai dari pemilihan data hingga metode penelitian yang digunakan, serta hasil pembahasan dan kesimpulan penelitian.

3. HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan berupa hasil *Clustering* melalui proses CRISP-DM dengan menggunakan algoritma K-Means yang didukung oleh tools Google Colab dengan bahasa pemrograman Python.

3.1 Bussiness Understanding

Melalui pemahaman yang komprehensif, penelitian ini berusaha memberikan kontribusi dalam upaya pencegahan dan penanganan stunting, sehingga mendorong peningkatan kesejahteraan balita di wilayah Kabupaten Dompu. Penelitian ini dilakukan dengan menggunakan metode *Clustering* dan algoritma K-Means yang dikombinasikan dengan *Principal Component Analysis* (PCA). Tujuan dalam penerapan data mining ini adalah melakukan *Clustering* Kabupaten Dompu di provinsi Nusa Tenggara Barat dengan mengelompokkan kawasan menjadi zona yang memiliki karakteristik berdasarkan faktor penyebab tingginya tingkat prevalensi stunting pada balita menggunakan algoritma K-Means yang dikombinasikan dengan PCA. Hasil pengelompokkan zona kemudian diidentifikasi sebagai kawasan yang menjadi referensi bagi pemerintah dalam mengambil kebijakan yang sesuai terkait penurunan prevalensi stunting di Kabupaten Dompu.

3.2 Data Understanding

Data dalam penelitian ini merupakan data sekunder yang diperoleh dari Dinas Kesehatan Kabupaten Dompu yang berupa data profil Kesehatan di tiap Kecamatan/Desa di Kabupaten Dompu tahun 2023. Variabel yang digunakan ditunjukkan pada Tabel 2

Tabel 2. Variabel Penelitian

Kode Variabel	Keterangan
TTD	Jumlah Ibu hamil penerima Tablet Tambah Darah
IMD	Jumlah Bayi menerima Inisiasi Menyusui Dini
ASIE	Jumlah Bayi Usia dibawah 6 Bulan Penerima ASI Eksklusif
VITA	Jumlah Bayi Menerima Vitamin A Usia 6-11 bulan
IDL	Jumlah Bayi Menerima Imunisasi Dasar Lengkap
ATFSL	Jumlah Keluarga dengan Akses Terhadap Fasilitas Sanitasi yang Layak

Berdasarkan Tabel 2, data yang digunakan menggunakan format .xlsx (Excel). Atribut yang terdapat dalam data tersebut terdiri dari 8 atribut yaitu Nomor, Kecamatan, Desa, TTD, IMD, ASIE, VITA, IDL, ATFSL.

3.3 Data Preparation

Pada tahap ini hal pertama yang dilakukan adalah memilih atribut yang akan digunakan untuk proses modelling dengan cara menghapus data yang tidak diperlukan. Atribut yang dihapus adalah Nomor. Data dianggap tidak diperlukan karena tidak berhubungan dengan penelitian atau atribut tersebut tidak berpengaruh terhadap proses analisis data.

Setelah memilih atribut yang akan digunakan untuk proses modelling tahap selanjutnya yaitu mengatasi missing value seperti menghapus kolom yang hilang dan menghapus data yang tidak relevan. Hal ini dilakukan agar proses modelling mendapatkan hasil yang baik

```
#Mencari missing value
print(data.isna().sum())

Kecamatan    0
Desa          0
TTD           0
IMD           0
ASIE          0
VITA          0
IDL           0
ATFSL        0
dtype: int64
```

Gambar 2. Mencari *Missing Value*

Berdasarkan gambar 2, tidak ditemukan adanya *missing value* yang ditemukan pada data. Langkah berikutnya adalah melakukan pengecekan terhadap redundansi data, yaitu data yang terindikasi memiliki duplikat atau kemiripan, sehingga apabila data yang dihasilkan lebih konsisten, dan akurat dapat mempengaruhi analisis selanjutnya.

```
#Mencari data yang terduplikat
data.duplicated().sum()

0
```

Gambar 3. Mencari data yang terduplikat

Berdasarkan gambar 3, tidak ditemukan adanya data yang terduplikat. Setelah pengecekan data duplikat telah selesai dilakukan. Langkah selanjutnya yaitu tahap transformasi data. Pada tahap ini dilakukan perubahan data kompleks menjadi lebih mudah diolah menggunakan fungsi penskalaan. Standarisasi data dilakukan jika data memiliki banyak variabel dengan rentang nilai yang cukup jauh. Tujuan dari standarisasi data yaitu agar data dengan variabel berbeda memiliki rentang nilai yang sama.

3.4 Modelling

3.4.1 Principal Component Analysis (PCA)

Penentuan komponen utama dapat dicapai dengan menggunakan *Variance Proportion Cumulative* (VPC) yang dapat menjelaskan jumlah varians total data hingga 70-80%.

Input:	num_components: jumlah komponen utama yang akan dipertahankan
Output:	pca_data: dataset yang diproyeksikan ke komponen utama pertama <code>`num_components`</code>
	explained_variance_: nilai eigen varians
	explained_variance_ratio_: proporsi varians yang dijelaskan oleh setiap komponen utama
	cumulative: jumlah kumulatif dari rasio varians yang dijelaskan
	%cumulative: persentase dari kumulatif varians yang dijelaskan

Gambar 4. Pseudocode PCA

Tabel 3. Eigenvalue berdasarkan hasil PCA

Variabel	Eigen Value	Total Variance (%)	Cumulative (%)
PC1	3.6841	58.7	58.7
PC2	0.8967	14.28	72.9
PC3	0.6347	10.11	83.1
PC4	0.4770	7.6	90.7
PC5	0.3560	5.6	96.3

Pada Tabel 3 menunjukkan dua variabel baru (faktor 1 dan faktor 2) atau komponen utama dengan nilai *Eigenvalue* tertinggi. Komponen utama pertama memiliki nilai eigen sebesar 3.68 yang mampu menjelaskan 58.7% variansi, komponen utama kedua memiliki nilai eigen sebesar 0.89 yang mampu menjelaskan 14.28% variansi.

Penciri faktor dapat dilihat dari variabel yang memiliki hubungan atau korelasi tinggi pada setiap faktor. Nilai variabel yang berupa komponen dari hasil PCA dengan korelasi tinggi memiliki batas nilai *cut off point* > 0.38.

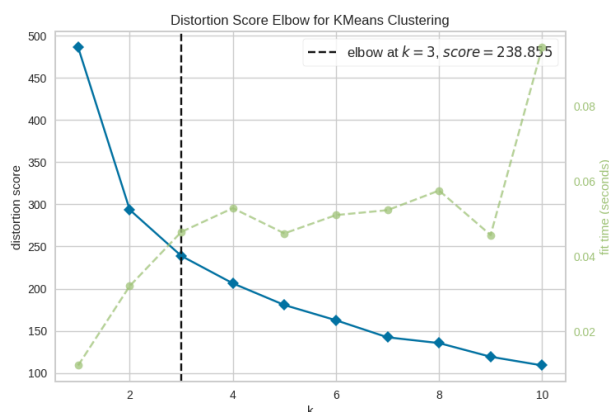
Tabel 4. Komponen hasil PCA dari setiap variabel

Variabel	Faktor 1	Faktor 2
TTD	0.43146268	-0.32958504
IMD	0.40762947	-0.13664326
ASIE	0.26261685	0.87810146
VITA	0.38263353	0.28288958
IDL	0.47064441	0.14629170
ATFSL	0.45912524	0.01457312

Pada Tabel 4 variabel penciri setiap *cluster* yang terbentuk adalah faktor 1 yang terdiri dari TTD, IMD, IDL dan ATFSL yang berkorelasi positif pada faktor, sedangkan variabel penciri pada faktor 2 terdiri dari ASIE.

3.4.2 Clustering

Untuk dapat melakukan pengelompokan data dengan K-Means perlu menentukan jumlah *Cluster* yang diperlukan. Optimalisasi penentuan jumlah *Cluster* dapat menggunakan metode *Elbow*. Titik yang menunjukkan *Cluster* optimal dalam metode ini akan membentuk siku dalam sebuah grafik

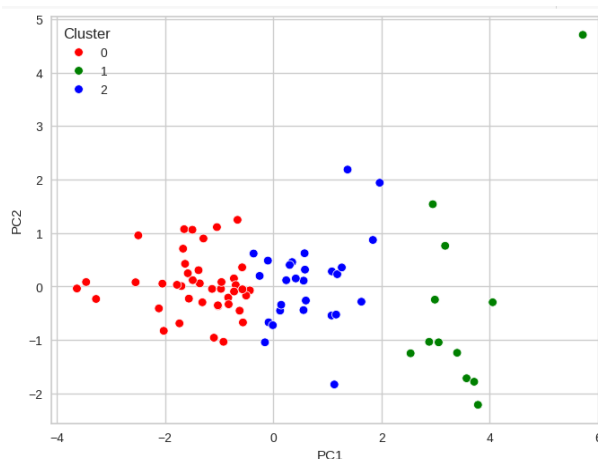


Gambar 5. Hasil K Optimal dengan metode *Elbow*

Berdasarkan Gambar 5 yang menunjukkan proses pencarian nilai K optimal, didapatkan bahwa jumlah kluster yang optimal dalam analisis ini adalah K=3

```
#Pilih K centroid awal secara acak dari data.
Repeat:
    Tetapkan setiap titik data ke kluster dengan centroid terdekat
    Hitung ulang posisi centroid untuk setiap kluster
Until :
    Posisi centroid tidak berubah
```

Gambar 6. Pseudocode K-Means Clustering



Gambar 7. Clustering K=3

Pada Gambar 7 menunjukkan hasil *cluster* dari data Kabupaten Provinsi Dompu yang membentuk 3 kawasan yaitu Cluster 0 (merah) Zona Tipe 1, Cluster 1 (Hijau) Zona Tipe 2, dan Cluster 2 (Biru) Zona Tipe 3. Tujuan dari pembentukan zona adalah untuk memastikan efektivitas kebijakan yang relevan dalam upaya menurunkan prevalensi stunting.

Tabel 5. Nilai *means* setiap *Cluster* K=3 untuk identifikasi tingkat stunting

Variabel	Nilai <i>Means</i> Setiap <i>Cluster</i> K=3		
	<i>Cluster</i> 0	<i>Cluster</i> 1	<i>Cluster</i> 2
Faktor 1	-1.405396	3.484042	0.637708
Faktor 2	0.036956	-0.316586	0.083217

Pada Tabel 5, angka negatif menunjukkan bahwa data berada di bawah rata-rata total dan berkorelasi positif, sedangkan nilai positif menunjukkan data berada di atas rata-rata total dan berkorelasi negatif.

Variabel penciri *Cluster* adalah variabel utama yang diidentifikasi dari analisis PCA. Dari tabel 4.4, dapat disusun tipologi untuk setiap *Cluster* dengan membandingkan nilai rata-rata antar *Cluster* pada variabel penciri tersebut.

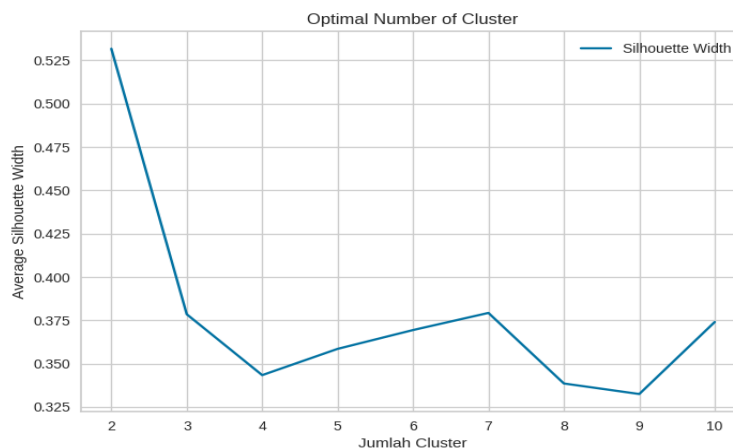
Tabel 6. Interpretasi perbandingan nilai *means* K=3 setiap *Cluster*

Variabel	Zona		
	<i>Cluster</i> 0	<i>Cluster</i> 1	<i>Cluster</i> 2
Faktor 1	Kurang	Baik	Cukup
Faktor 2	Cukup	Kurang	Baik

Hasil dari Tabel 6 menunjukkan karakteristik pengelompokan objek atau *Cluster* digunakan untuk setiap kawasan yang membentuk Zona.

3.5 Evaluation

Langkah selanjutnya yaitu melakukan tahap evaluasi untuk mengukur hasil atau kualitas dari *Cluster* dan model yang dibentuk. Teknik yang digunakan yaitu *Silhouette Coefficient*.



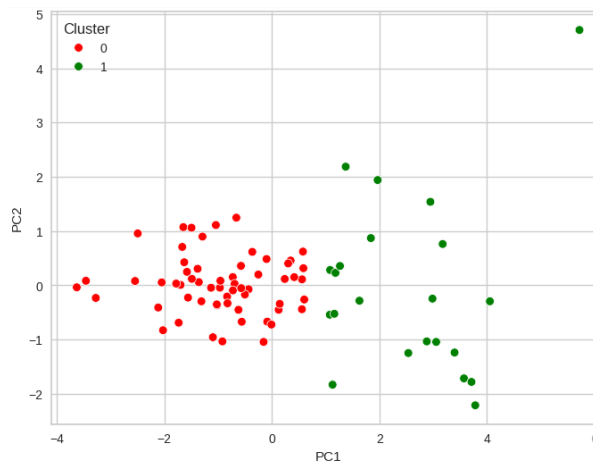
Gambar 8. Grafik *silhouette*

Pada gambar 8 dalam penentuan jumlah *Cluster* berdasarkan metode *silhouette coefficient*, *Cluster* optimal berada pada titik K=2.

Tabel 7. Hasil evaluasi nilai K pada K-Means

K	DBI
2	0.779
3	0.856
4	0.927

Pada Davies Bouldin Index (DBI), pengelompokan data dinyatakan optimal atau semakin baik jika memiliki index yang lebih rendah atau mendekati angka 0. Berdasarkan Tabel 7, nilai index yang lebih rendah terdapat pada K=2.



Gambar 9. Clustering K=2

Pada Gambar 9 menunjukkan hasil *cluster* dari data Kabupaten Provinsi Dompus yang membentuk 2 kawasan yaitu Cluster 0 (merah) Zona Tipe 1, dan Cluster 2 (Hijau) Zona Tipe 2.

Tabel 8. Nilai *means* setiap *Cluster* K=2 untuk identifikasi tingkat stunting

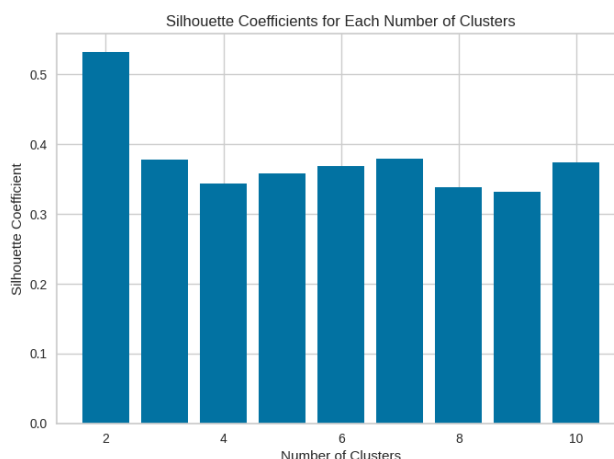
Variabel	Nilai Means Setiap Cluster K=2	
	Cluster 0	Cluster 1
Faktor 1	-0.940954	2.523466
Faktor 2	0.018772	-0.050343

Pada Tabel 8, angka negatif menunjukkan bahwa data berada di bawah rata-rata total dan berkorelasi negatif, sedangkan nilai positif menunjukkan data berada di atas rata-rata total dan berkorelasi positif.

Variabel penciri *Cluster* adalah variabel utama yang diidentifikasi dari analisis PCA. Dari tabel 8, dapat disusun tipologi untuk setiap *Cluster* dengan membandingkan nilai rata-rata antar *Cluster* pada variabel penciri tersebut yang ditunjukkan pada Tabel 9.

Tabel 9. Interpretasi perbandingan nilai *means* K=2 setiap *Cluster*

Variabel	Zona	
	Cluster 0	Cluster 1
Faktor 1	Kurang	Baik
Faktor 2	Baik	Kurang



Gambar 10. Silhouette Coefficients

Pada Gambar 10 menunjukkan bahwa indeks silhouette coefficient dengan jumlah *Cluster* K=2 adalah 0.53 sehingga dapat dikatakan bahwa model segmentasi tersebut memiliki struktur yang baik atau sedang (*medium structure*) karena berada pada rentang 0.51-0.70.

3.6 Deployment

Berdasarkan variabel penciri yang terbentuk pada faktor 1 dan 2 dalam Tabel 8 dan Tabel 9, karakteristik setiap Zona dapat diidentifikasi dengan mengetahui ciri khas yang membedakan satu Zona dari Zona lainnya.

3.6.1 Tipe Zona berdasarkan hasil pembagian 3 Cluster

Tabel 10. Hasil *Cluster* Kecamatan/Desa berdasarkan Tipe Zona dengan jumlah $K=3$

	Kecamatan	Desa
Zona Tipe 1	Dompu	Katua, Mangge Nae
	Woja	Riwo, Mumbu, Rababaka, Serakapi
	Pajo	Lune, Woko, Ranggo, Lepadi, Temba Lae
	Hu'u	Adu, Sawe, Jala, Marada
	Kempo	Soro, Konte, So Nggajah, Tolo Kalo, Doro Kobo, Soro Barat
	Kilo	Mbuju, Taropo, Kramat, K I W U
	Manggelewa Pekat	Kwangko, Nangatumpu, Soriutu, Lanci Jaya, Nusa Jaya, Tanju, Teka Sire, Anamina Doropeti, Beringin Jaya, Sorinomo, Nangakara, Sori Tatanga, Tambora, Nangamiro, Karombo
Zona Tipe 2	Dompu	Kandai Satu, Karijawa, Potu, Bada, Bali, Dorotangga, Oo
	Woja	Matua, Montabaru, Kandai Dua, Simpasai
	Pekat	Pekat
Zona Tipe 3	Dompu	Mangge Asi, Sorisakolo, Mbawi, Dorebara, Karamabura, Kareke
	Woja	Madaprama, Bara, Nowa, Wawonduru, Saneo, Bakajaya
	Hu'u	Hu'u, Daha, Rasabou, Cempi Jaya
	Kempo	Ta'a, Kempo
	Kilo	Malaju, Lasi
Manggelewa Pekat	Banggo, Doromelo, Suka Damai, Kampasi Meci Kadindi, Calabai, Kadindi Barat	

a. Zona Tipe 1

Cluster ini bercirikan sebagai kawasan yang memiliki jumlah ibu hamil penerima Tablet Tambah Darah (TTD), jumlah bayi yang menerima Inisiasi Menyusui Dini (IMD), jumlah bayi menerima Imunisasi Dasar Lengkap (IDL), dan jumlah keluarga dengan akses terhadap fasilitas sanitasi yang layak yang masih kurang, sementara jumlah bayi usia dibawah 6 bulan yang menerima ASI eksklusif cukup. Dengan demikian dapat diinterpretasikan bahwa *Cluster* 1 merupakan kawasan yang dijadikan sebagai Zona Tipe 1.

b. Zona Tipe 2

Cluster ini bercirikan sebagai kawasan yang memiliki jumlah ibu hamil penerima Tablet Tambah Darah (TTD), jumlah bayi yang menerima Inisiasi Menyusui Dini (IMD), jumlah bayi menerima Imunisasi Dasar Lengkap (IDL), dan jumlah keluarga dengan akses terhadap fasilitas sanitasi yang layak dapat dikatakan baik, sementara jumlah bayi usia dibawah 6 bulan yang menerima ASI eksklusif masih kurang. Dengan demikian dapat diinterpretasikan bahwa *Cluster* 2 merupakan kawasan yang dijadikan sebagai Zona Tipe 2.

c. Zona Tipe 3

Cluster ini bercirikan sebagai kawasan yang memiliki jumlah ibu hamil penerima Tablet Tambah Darah (TTD), jumlah bayi yang menerima Inisiasi Menyusui Dini (IMD), jumlah bayi menerima Imunisasi Dasar Lengkap (IDL), dan jumlah keluarga dengan akses terhadap fasilitas sanitasi yang layak dapat dikatakan cukup, sementara jumlah bayi usia dibawah 6 bulan yang menerima ASI eksklusif adalah baik. Dengan demikian dapat diinterpretasikan bahwa *Cluster* 3 merupakan kawasan yang dijadikan sebagai Zona Tipe 3.

3.6.2 Tipe Zona berdasarkan hasil pembagian 2 Cluster

Tabel 11. Hasil *Cluster* Kecamatan/Desa berdasarkan Tipe Zona dengan jumlah $K=2$

	Kecamatan	Desa
Zona Tipe 1	Dompu	Mbawi, Katua, Karamabura, Kareke, Mangge Nae
	Woja	Riwo, Madaprama, Mumbu, Rababaka, Serakapi
	Pajo	Jambu, Lune, Woko, Ranggo, Lepadi, Temba Lae
	Hu'u	Hu'u, Daha, Rasabou, Cempi Jaya, Adu, Sawe, Jala, Marada
	Kempo	Ta'a, Kempo, Soro, Konte, So Nggajah, Tolo Kalo, Doro Kobo, Soro Barat
	Kilo	Mbuju, Taropo, Kramat, Malaju, Lasi, K I W U
	Manggelewa	Kwangko, Nangatumpu, Banggo, Soriutu, Doromelo, Lanci Jaya, Nusa Jaya, Suka Damai, Tanju, Kampasi Meci, Teka Sire, Anamina
	Pekat	Doropeti, Beringin Jaya, Sorinomo, Nangakara, Sori Tatanga, Tambora, Nangamiro, Calabai, Karombo
Zona Tipe 2	Dompu	Kandai Satu, Karijawa, Potu, Bada, Bali, Dorotangga, Mangge Asi, Sorisakolo, Dorebara, Oo
	Woja	Bara, Nowa, Wawonduru, Matua, Montabaru, Kandai Dua, Simpasai, Saneo, Bakajaya
	Pekat	Pekat, Kadindi, Kadindi Barat

a. Zona Tipe 1

Cluster ini bercirikan sebagai kawasan yang memiliki jumlah ibu hamil penerima Tablet Tambah Darah (TTD), jumlah bayi yang menerima Inisiasi Menyusui Dini (IMD), jumlah bayi menerima Imunisasi Dasar Lengkap (IDL), dan jumlah keluarga dengan akses terhadap fasilitas sanitasi yang layak masih kurang, sementara jumlah bayi usia dibawah 6 bulan yang menerima ASI eksklusif dapat dikatakan baik. Dengan demikian dapat diinterpretasikan bahwa *Cluster 1* merupakan kawasan yang dijadikan sebagai Zona Tipe 1.

b. Zona Tipe 2

Cluster ini bercirikan sebagai kawasan yang memiliki jumlah ibu hamil penerima Tablet Tambah Darah (TTD), jumlah bayi yang menerima Inisiasi Menyusui Dini (IMD), jumlah bayi menerima Imunisasi Dasar Lengkap (IDL), dan jumlah keluarga dengan akses terhadap fasilitas sanitasi yang layak dapat dikatakan baik, sementara jumlah bayi usia dibawah 6 bulan yang menerima ASI eksklusif masih kurang. Dengan demikian dapat diinterpretasikan bahwa *Cluster 2* merupakan kawasan yang dijadikan sebagai Zona Tipe 2

4. KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa metode analisis K-Means dengan jumlah *Cluster* optimal yang dihasilkan menggunakan metode *Elbow* adalah sebanyak 3 *Cluster* yaitu terdiri dari Zona Tipe 1 yang terdiri dari 42 Desa, Zona Tipe 2 yang terdiri dari 12 Desa dan Zona Tipe 3 yang terdiri dari 27 Desa, sedangkan hasil *Clustering* menggunakan metode *silhouette* menunjukkan 2 *Cluster* yaitu Zona Tipe 1 yang terdiri dari 54 Desa, dan Zona Tipe 2 yang terdiri dari 22 Desa dengan *Score silhouette* sebesar 0.53 menunjukkan bahwa model *Cluster* pada data memiliki struktur yang cukup baik (*medium structure*). Penentuan faktor dominan dapat diperoleh dari hasil PCA yang berupa komponen utama. Hasil yang diperoleh dari perhitungan PCA adalah dua variabel baru (faktor 1 dan faktor 2) atau komponen utama dengan nilai *Eigen value* tertinggi. Komponen utama pertama memiliki nilai *eigen* sebesar 3.68 yang mampu menjelaskan 58.7% variansi yang terdiri dari variabel jumlah ibu hamil penerima Tablet Tambah Darah (TTD), jumlah bayi yang menerima Inisiasi Menyusui Dini (IMD), jumlah bayi menerima Imunisasi Dasar Lengkap (IDL), dan jumlah keluarga dengan Akses Terhadap Fasilitas Sanitasi yang Layak (ATFSL), komponen utama kedua memiliki nilai *eigen* sebesar 0.89 yang mampu menjelaskan 14.28% variansi berupa variabel jumlah bayi usia dibawah 6 bulan yang menerima ASI eksklusif (ASIE). Oleh karena itu, nilai kumulatif variansi mencapai 72.9%, menunjukkan bahwa terdapat 2 komponen utama yaitu Faktor 1 yang terdiri dari TTD, IMD, IDL dan ATFSL, dan Faktor 2 yang terdiri dari ASIE yang dapat menjadi karakteristik kawasan suatu Zona yang mempengaruhi prevalensi *stunting* di Kabupaten Dompur.

REFERENCES

- [1] M. V. Moxin, R. Muslim, and S. N. Ishak, "Pencegahan Stunting Di Wilayah Ternate Melalui Perubahan Perilaku Wanita Hamil," *Jurnal Ilmiah Serambi Sehat*, vol. 15, no. 1, 2022.
- [2] D. R. Sari, "Metode Principal Component Analysis (PCA) Sebagai Penanganan Asumsi Multikolinearitas (Studi Kasus: Data Produksi Tapioka) Principal Component Analysis (PCA) Method to Handle Multicollinearity Assumptions (Case Study: Tapioca Production Data)," *Parameter: Jurnal Matematika, Statistika dan Terapannya*, vol. 2, no. 2, 2023, doi: 10.30598/parameterv2i02pp115-124.
- [3] H. Shen and Z. Duan, "Application research of clustering algorithm based on K-means in data mining," 2020 International Conference on Computer Information and Big Data Applications (CIBDA), 2020, doi: 10.1109/CIBDA50819.2020.00023.
- [4] S. S. Dewanti and A. J. B. Pramono, "Dominant factors that determine college students completing studies in mathematics education study programs," *REID (Research and Evaluation in Education)*, vol. 9, no. 1, 2023, doi: 10.21831/reid.v9i1.51081.
- [5] K. Hilmi Izzuddin and A. W. Wijayanto, "Pemodelan Clustering Ward, K-Means, DIANA, dan PAM dengan PCA untuk Karakterisasi Kemiskinan Indonesia Tahun 2021," *Komputika: Jurnal Sistem Komputer*, vol. 13, no. 1, 2024, doi: 10.34010/komputika.v13i1.10803.
- [6] S. Surono, K. W. Goh, C. W. Onn, and F. Marestiani, "Developing an optimized recurrent neural network model for air quality prediction using K-means clustering and PCA dimension reduction," *International Journal of Innovative Research and Scientific Studies*, vol. 6, no. 2, 2023, doi: 10.53894/ijirss.v6i2.1427.
- [7] I. A. Rosyada and D. T. Utari, "Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering," *Jambura Journal of Probability and Statistics*, vol. 5, no. 1, 2024. doi: 10.34312/jjps.v5i1.18733.
- [8] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient," *Journal of Applied Informatics and Computing (JAIC)*, Vol.5, No.2, 2021, doi: 10.30871/jaic.v5i1.3237.
- [9] A. Iskandar, R. Aldy Al Hafizh Harahap, and A. Gilang Ramadhan, "Data Mining Customer Clustering Using K-Means Method," *International Journal of Economic Literature*, vol. 2, no. 3, pp. 927–944, 2024.
- [10] D. A. Awaliyah, B. Prasetyo, R. Muzayanah, and A. D. Lestari, "Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm," *Scientific Journal of Informatics*, vol. 11, no. 2, 2024, doi: 10.15294/sji.v11i2.6137.
- [11] S. Suraya, M. Sholeh, and U. Lestari, "Evaluation of Data Clustering Accuracy using K-Means Algorithm," *International Journal of Multidisciplinary Approach Research and Science*, vol. 2, no. 01, pp. 385–396, 2023, doi: 10.59653/ijmars.v2i01.504.
- [12] E. Santri and D. Brahma, "Journal of Artificial Intelligence and Engineering Applications Analysis Of Poverty Level Mapping In Riau Province Using The K-Means Method," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 3, no. 2, 2024, doi: 10.59934/jaiea.v3i2.414.
- [13] R. Yuniar Putri Sayda et al., "Pengetahuan Ibu Hamil tentang Konsumsi Tablet Tambah Darah sebagai Upaya Pencegahan

- Stunting di Surabaya Timur,” *Jurnal Farmasi Komunitas* vol. 11, no. 1, 2024, [Online]. Available: <https://orcid.org/0000-0002-6023-9326>.
- [14] S. S. Lintang and F. Azkiya, “Hubungan Inisiasi Menyusu Dini (Imd) Dengan Kejadian Stunting Pada Bayi Usia 0-24 Bulan Di Puskesmas Kramatwatu Tahun 2021,” *Journal of Midwifery*, vol. 10, no. 2, pp. 155–160, 2022, doi: 10.37676/jm.v10i2.3274.
- [15] D. P. Windasari, I. Syam, and L. S. Kamal, “Faktor hubungan dengan kejadian stunting di Puskesmas Tamalate Kota Makassar,” *AcTion: Aceh Nutrition Journal*, vol. 5, no. 1, p. 27, 2020, doi: 10.30867/action.v5i1.193.
- [16] A. Yuniarti and A. Setiari, “Konsumsi Vitamin A untuk Mencegah Kejadian Stunting : Systematic Review,” *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, vol. 7, no. 3, pp. 588–595, 2024, doi: 10.56338/mppki.v7i3.4784.
- [17] E. Noorhasanah, N. Isna Tauhidah, M. Chalida Putri, P. S. Studi, and K. Ners Fakultas Keperawatan dan Ilmu Kesehatan Universitas Muhammadiyah Banjarmasin, “Faktor-Faktor Yang Berhubungan Dengan Kejadian Stunting Pada Balita di Wilayah Kerja Puskesmas Tatah Makmur Kabupaten Banjar (Factors Related to Stunting Evens In Children In The Working Area of Puskesmas Tatah Makmur)” *Journal of Midwifery and Reproduction*, vol. 4, no. 1, 2020, doi: 10.35747/jmr.v4i1.559.
- [18] M. Adzura, F. Fathmawati, Y. Yulia, J. K. Lingkungan, P. Kesehatan, and K. Pontianak, “Hubungan Sanitasi, Air Bersih Dan Mencuci Tangan Dengan Kejadian Stunting Pada Balita Di Indonesia Relationship of Sanitation, Clean Water, and Handwashing with Stunting in Children in Indonesia.” *Jurnal Sulolipu : Media Komunikasi Sivitas Akademika dan Masyarakat.*, vol. 21 no.1 2021, doi: 10.32382/sulolipu.v21i1.2098.
- [19] N. A. Maori, “Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering,” *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 1, 2023, doi: 10.24176/simet.v14i2.9630.
- [20] N. Nugroho and F. D. Adhinata, “Penggunaan Metode K-Means dan K-Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa,” *TEKNIKA: Jurnal Teknologi Informasi dan Komunikasi*, vol. 11, no. 3, pp. 170–179, 2022, doi: 10.34148/teknika.v11i3.502.