BAB II
TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna (Walpole, 1995). Statistika deskriptif bertujuan untuk menyajikan informasi yang terkait dengan data yang dimiliki tanpa melakukan interpretasi atau kesimpulan. Dalam proses ini, penyusunan tabel, diagram, grafik, dan pengukuran lainnya merupakan bagian dari kategori statistika deskriptif yang membantu dalam menggambarkan secara detail karakteristik data.

2.1.1 Rata-rata (*Mean*)

Rata-rata hitung (*mean*) merupakan salah satu pengukuran yang dapat digunakan untuk memberikan gambaran yang lebih jelas dan singkat tentang sekumpulan data dengan melihat pusat suatu data. Rata-rata populasi dilambangkan dengan μ, sedangkan untuk rata-rata sampel dilambangkan dengan $\bar{x}$ (x bar). Untuk dapat menentukan *mean* maka dapat dilakukan dengan cara menjumlahkan seluruh nilai data kemudian dibagi banyaknya data yang ada. Berikut adalah rumus untuk menghitung rata-rata hitung:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.1}$$

dimana,

$\bar{x}$: Nilai rata-rata (*mean*)

x_i : Nilai data ke- i

n : Banyak data

2.1.2 Standar Deviasi

Standar deviasi atau simpangan baku merupakan nilai statistik yang menunjukkan tingkat (derajat) variasi, digunakan dalam menentukan persebaran suatu data pada sampel yang ada dan untuk melihat seberapa dekat data-data yang

ada dengan mean atau rata-rata nilai sampel. Berikut adalah rumus untuk menghitung standar deviasi:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad (2.2)$$

dimana,

S : Standar deviasi

x_i : Nilai data ke- i

$\bar{x}$: Nilai rata-rata

n : Banyak data

2.1.3 Nilai Maksimum dan Minimum

Nilai maksimum dan minimum didefinisikan sebagai berikut:

1. Nilai Maksimum

   Nilai maksimum adalah nilai tertinggi atau puncak dari sekumpulan data. Ini adalah nilai atau titik yang lebih besar daripada semua nilai lain dalam dataset tersebut. Dalam analisis statistik, nilai maksimum dapat memberikan informasi tentang batas atas dari data.

2. Nilai Minimum

   Nilai minimum adalah nilai terendah atau titik terendah dari sekumpulan data. Ini adalah nilai atau titik yang lebih kecil daripada semua nilai lain dalam dataset tersebut. Dalam analisis statistik, nilai minimum dapat memberikan informasi tentang batas bawah dari data.

2.2 Uji Multikolinearitas

Multikolinearitas merupakan sebuah kondisi dimana antara satu pasang variabel independen atau lebih memiliki korelasi yang kuat (Best & Wolf dalam Kusumawardhani, dkk, 2021). Dalam analisis pengelompokan, adanya korelasi antar variabel bisa menyulitkan dalam menentukan pengaruh dari setiap masing-masing variabel. Oleh karena itu, sebelum melakukan pengelompokan penting untuk melakukan uji multikolinearitas. Salah satu cara yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah dengan menggunakan nilai *Variance Inflation Factor* (VIF). Menurut Best & Wolf dalam Kusumawardhani, dkk (2021),

nilai VIF digunakan untuk mengukur sejauh mana variabel independen ke-j bergantung pada kumpulan variabel independen yang lain. Persamaan VIF dapat dilihat pada persamaan berikut:

$$VIF_j = \frac{1}{1-R_j^2} \tag{2.3}$$

dimana,

R_j^2 : Koefisien determinasi variabel ke-j

m : Banyak variabel yang diamati

j : 1, 2, 3,..., m

Koefisien determinasi dapat diperoleh melalui persamaan berikut:

$$R_j^2 = \frac{b_1 \sum x_1 x_j + b_2 \sum x_2 x_j + \cdots + b_m \sum x_m x_j}{\sum x_j^2} \tag{2.4}$$

dengan $\sum x_j^2 = \sum X_j^2 - \frac{\left(\sum X_j\right)^2}{n}$

$$\sum x_m x_j = \sum X_m X_j - \frac{(\sum X_m)(\sum X_j)}{n}$$

dimana,

$\sum X_m$: Jumlah data dari X pada variabel ke-m

$\sum X_j$: Jumlah data dari X pada variabel ke-j

$\sum X_j^2$: Jumlah data dari X^2 pada variabel ke-j

$\sum X_m X_j$: Jumlah data dari $X_m X_j$

n : Banyak data

b_m : Koefisien regresi masing-masing variabel

Multikolinearitas juga dapat diidentifikasi dengan menghitung nilai *tolerance* dengan menggunakan persamaan berikut:

$$Tolerance = \frac{1}{VIF} \tag{2.5}$$

$$= 1 - R_j^2$$

Suatu model dikatakan memiliki multikolinearitas jika nilai VIF > 10 atau $Tolerance \leq 0{,}1$.

2.3 Analisis Pengelompokan

Analisis pengelompokan adalah suatu metode statistik yang mengelompokan sampel berdasarkan kesamaan karakteristik. Tujuan dari analisis pengelompokan adalah untuk membentuk kelompok yang berbeda dengan tingkat kesamaan tinggi di antara anggota kelompok dan tingkat perbedaan tinggi di antara kelompok yang berbeda. Pengelompokan ini didasarkan pada variabel-variabel tertentu yang ingin diteliti.

Analisis pengelompokan harus memenuhi asumsi berikut:

1. Sampel yang diambil harus benar-benar bisa mewakili populasi.
2. Multikolinieritas yaitu korelasi antar objek. Sebaiknya tidak terdapat korelasi antar objek, bila ada maka besar multikolinieritas tidaklah tinggi yaitu (< 0,5).

Sebelum melakukan analisis pengelompokan, perlu memperhatikan data yang akan digunakan. Jika terdapat perbedaan nilai yang besar antara variabel, misalnya ada yang dalam satuan juta dan ada yang satuan puluhan atau bahkan lebih kecil, perbedaan tersebut akan menyebabkan perhitungan jarak menjadi tidak valid. Oleh karena itu, data perlu ditransformasi menggunakan uji *z-score*. Tujuan dari analisis pengelompokan adalah untuk mengklasifikasikan objek ke dalam kelompok yang relatif homogen, objek-objek di dalam satu kelompok memiliki kesamaan yang lebih tinggi dibandingkan dengan objek-objek pada kelompok lain.

Untuk menilai tingkat kemiripan antara dua objek, diperlukan definisi tentang ukuran kemiripan tersebut. Tujuannya adalah untuk membuat matriks *proximity* yang bersifat persegi dan simetri dengan jumlah objek yang sama pada baris dan kolom. Matriks ini mencerminkan tingkat kemiripan atau perbedaan antara objek-objek tersebut. Untuk mengukur kemiripan antar data digunakan pengukuran jarak (*distance*). Semakin besar jarak antara dua buah data maka semakin besar perbedaan antara dua data. Pengukuran jarak ada bermacam-macam namun yang paling sering digunakan pada pengelompokan adalah jarak *Euclidean* dengan rumus berikut:

$$d_{(i,k)} = \sqrt{\sum_{j=1}^{m}(x_{ij} - x_{kj})^2} \tag{2.6}$$

dimana,

$d_{(i,k)}$: Jarak antara data ke-*i* dengan data ke-*k*

m : Banyak variabel yang diamati

x_{ij} : Data ke-*i* pada variabel ke-*j*

x_{kj} : Data ke-*k* pada variabel ke-*j*

Secara umum metode analisis pengelompokan terbagi menjadi dua metode yaitu metode hierarki dan metode non-hierarki.

2.3.1 Metode Hierarki

Metode hierarki digunakan ketika jumlah kelompok yang akan dibentuk belum diketahui. Pada metode hierarki *agglomerative,* proses dimulai dengan setiap data berada dalam kelompok terpisah sebanyak *n* kelompok. Kemudian, kelompok-kelompok tersebut digabungkan dengan menggabungkan dua data yang paling dekat satu sama lain, lalu mengevaluasi kembali kedekatan antara *n-1* kelompok yang baru terbentuk. Proses penggabungan ini berlanjut dengan menggabungkan dua data terdekat dan terus berlanjut hingga akhirnya terbentuk satu kelompok tunggal yang berisi seluruh data (Nugroho, 2008). Metode yang termasuk dalam metode *agglomerative* yaitu:

1. Pautan tunggal (*Single linkage*)

Untuk menentukan jarak antar kelompok dengan menggunakan *Single Linkage*, maka dipilih jarak yang paling dekat atau aturan tetangga dekat (*nearest neighbour rule*). Berikut adalah persamaan metode *Single Linkage*:

$$D(A,B) = \min(d_{i,k}) \tag{2.7}$$

dimana,

$D(A,B)$: Jarak antara kelompok *A* dan *B*

$d_{(i,k)}$: Jarak antara data ke-*i* dengan data ke-*k*

2. Pautan lengkap (*Complete linkage*)

Metode *Complete Linkage* merupakan metode pembaruan matriks kemiripan yang berdasarkan pada jarak terjauh antar objeknya (*maksimum distance*). Berikut adalah persamaan metode C*omplete linkage*:

$$D(A,B) = \max(d_{i,k}) \tag{2.18}$$

dimana,

$D(A,B)$: Jarak antara kelompok *A* dan *B*

$d_{(i,k)}$: Jarak antara data ke-*i* dengan data ke-*k*

3. Pautan rata-rata (*Average linkage*)

Metode A*verage linkage* merupakan metode pembaruan matriks kemiripan yang berdasarkan pada jarak rata-rata antar objeknya (*average distance*). Metode ini bertujuan meminimumkan rataan jarak semua pasangan pengamatan dari dua kelompok yang digabung. Berikut adalah persamaan metode *Average linkage*:

$$D(A,B) = \frac{\Sigma_a \Sigma_b d_{i,k}}{n_A n_B} \tag{2.19}$$

dimana,

$D(A,B)$: Jarak antara kelompok *A* dan *B*

$d_{(i,k)}$: Jarak antara data ke-*i* dengan data ke-*k*

n_A : Jumlah data pada kelompok *A*

n_B : Jumlah data pada kelompok *B*

4. Metode Ward (*Ward's method*)

Metode ward adalah metode yang didasarkan pada hasil informasi yang minimum dari kenaikan pada jumlah kuadrat deviasi rataan kelompok. Proses berhenti pada kenaikan yang menyebabkan *Error Sum of Squares* (ESS) dari gabungan tiap kelompok yang mungkin. Nilai ESS digunakan sebagai fungsi objektif dan didefinisikan sebagai berikut:

$$ESS = \sum_{A=1}^{q} \left(\sum_{i=1}^{n_A} x_{iA}^2 - \frac{1}{n_A} \left(\sum_{i=1}^{n_A} x_{iA} \right)^2 \right) \tag{2.10}$$

dimana,

x_{iA} : Nilai data ke-*i* pada kelompok ke-*A*

q : Jumlah kelompok tiap *stage*

n_A : Jumlah data ke-*i* pada kelompok ke-*A*

5. Metode *Centroid* (*Centroid Method*)

Pada metode ini, jarak antara dua kelompok adalah jarak di antara dua *centroid* kelompok-kelompok tersebut. *Centroid* adalah vektor rata-rata jarak yang ada pada sebuah kelompok, yang didapat dengan melakukan rata-rata pada semua anggota suatu kelompok tertentu. Dengan metode ini, setiap terjadi pengelompokan baru, segera terjadi perhitungan ulang *centroid*, sampai terbentuk kelompok yang tetap. Jika $\bar{x}_A$ dan $\bar{x}_B$ adalah vektor rataan (*centroid*) kelompok *A* dan *B*, maka jarak antar dua kelompok didefinisikan sebagai $D(A,B) = d(\bar{x}_A, \bar{x}_B)$. *Centroid* kelompok baru yang terbentuk didapat dengan rumus :

$$\bar{x}_{AB} = \frac{n_A \bar{x}_A + n_A \bar{x}_B}{n_A + n_B} \tag{2.11}$$

dimana,

$\bar{x}_{AB}$: *Centroid* kelompok yang baru terbentuk

n_A : Banyaknya anggota kelompok *A*

n_B : Banyaknya anggota kelompok *B*

Untuk menentukan metode yang terbaik pada metode hierarki, digunakan uji korelasi *Cophenetic*. Korelasi *Cophenetic* didefinisikan sebagai koefisien korelasi linier antara jarak kofenetik yang diperoleh dari pohon dan jarak awal (ketidakmiripan) yang digunakan untuk membangun pohon. Ini merupakan ukuran seberapa setiap pohon itu mewakili perbedaan diantara pengamatan. Koefisien *Cophenetic* merupakan koefisien korelasi antara elemen-elemen asli matriks ketidakmiripan (*dissimilarity distance*) dan elemen-elemen yang dihasilkan oleh dendogram (*matrix cophenetic* berdasarkan ukuran jarak dan metode keterhubungan yang digunakan). Proses ini dimulai dengan menghitung matriks jarak asli (jarak *euclidean*) untuk membentuk matriks jarak antar data. Selanjutnya terapkan metode *single linkage*, *complete linkage*, *average linkage*, metode *ward* dan metode c*entroid*. Setelah itu, buat dendrogram untuk setiap metode tersebut dan hitung matriks jarak *cophenetic* dari masing-masing dendrogram. Kemudian, hitung korelasi *cophenetic* antara matriks jarak asli dan matriks jarak *cophenetic* untuk setiap metode menggunakan persamaan berikut:

$$r_{coph} = \frac{n\sum_{i<k}^{n} d_{ik}d_{c_{ik}} - \left(\sum_{i<k}^{n} d_{ik}\right)\left(\sum_{i<k}^{n} d_{c_{ik}}\right)}{\sqrt{\left[n\sum_{i<k}^{n} x_{ik}^2 - \left(\sum_{i<k}^{n} x_{ik}\right)^2\right]\left[n\sum_{i<k}^{n} x_{c_{ik}}^2 - \left(\sum_{i<k}^{n} x_{c_{ik}}\right)^2\right]}} \quad (2.12)$$

dimana,

r_{coph} : Koefisien korelasi *Cophenetic*

d_{ik} : Jarak asli (jarak *Euclidean*) data ke-*i* dengan data ke-*k*

$d_{c_{ik}}$: Jarak *Cophenetic* data ke-*i* dengan data ke-*k*

n : Banyak data

Nilai korelasi *Cophenetic* antara -1 sampai 1, yang semakin mendekati 1 berarti solusi yang dihasilkan semakin baik (Jain & Dubes, 1988). Metode dengan nilai korelasi mendekati 1 yang akan dipilih sebagai metode terbaik pada metode hierarki.

2.3.2 Metode Non Hierarki

Metode non-hierarki dimulai dengan menentukan terlebih dahulu jumlah kelompok yang diinginkan dan *centroid*. Metode ini biasa disebut pengelompokan *K-Means*. Pengelompokan *K-Means* dimulai dengan menetapkan terlebih dahulu jumlah kelompok yang diinginkan dan posisi titik pusat. Pada beberapa perangkat lunak, titik pusat awal dapat menggunakan *k* pengamatan pertama, tetapi ada juga yang menentukan titik pusat secara acak. Setelah itu, jarak antara setiap objek dengan setiap titik pusat dihitung. Objek kemudian ditempatkan ke dalam kelompok yang sesuai berdasarkan jarak terdekat dengan titik pusat. Selanjutnya hitung ulang titik pusat, proses ini diulang hingga tidak ada lagi pemindahan objek antar kelompok (Nugroho, 2008).

Pengelompokan menggunakan *K-Means* bertujuan untuk menggambarkan algoritma yang digunakan dalam menempatkan suatu objek ke dalam kelompok tertentu berdasarkan rata-rata terdekat. Metode *K-Means* kerap digunakan karena sederhana dan mudah diimplementasikan. Berikut adalah langkah-langkah algoritma *K-Means* :

1. Diberikan nilai *k* sebagai jumlah kelompok yang ingin dibentuk.
2. Tentukan *k centroid* (titik pusat) awal secara *random*.
3. Hitung jarak setiap data ke masing-masing titik pusat yaitu menggunakan jarak *Euclidean*.
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan pusatnya.
5. Tentukan posisi titik pusat baru dengan cara menghitung nilai rata-rata dari data-data yang ada pada kelompok yang sama.

$$C_{lj} = \frac{x_{1j}+x_{2j}+\cdots+x_{n_{lj}}}{n_l} \tag{2.13}$$

dimana,

C_{lj} : Nilai titik pusat baru pada data ke-*l* dan variabel ke-*j*

n_l : Banyak data pada kelompok ke-*l*

6. Hitung kembali jarak setiap data ke titik pusat yang baru terbentuk, begitu seterusnya hingga tidak ada lagi pemindahan objek antar kelompok. Jika tidak terjadi perpindahan objek antar kelompok, maka iterasi dihentikan.

2.4 Uji Validasi

Uji validasi digunakan untuk menganalisa hasil yang diperoleh dari algoritma pengelompokan. Pengujian ini bertujuan untuk memperoleh jumlah kelompok terbaik yang dapat menjelaskan keseluruhan struktur data. Jumlah kelompok terbaik dapat memberikan hasil pengelompokan yang optimal sehingga diperoleh keputusan yang tepat (Susilowati, dkk, 2020). Terdapat beberapa cara uji validasi diantaranya adalah *Silhouette Coefficient Index, Indeks Dunn, Indeks Connectivity*, *Indeks Calinski-Harabasz, Indeks Davies-Bouldin*, dan *Indeks Ball*. Pada penelitian ini akan digunakan uji *Silhouette Coefficient Index*. *Silhouette Coefficient Index* (SC) memiliki keunggulan karena hanya bergantung pada partisi sebenarnya dari objek, bukan pada algoritma pengelompokan yang digunakan untuk memperolehnya. SC dapat digunakan untuk meningkatkan hasil analisis pengelompokan atau untuk membandingkan hasil dari berbagai algoritma pengelompokan yang diterapkan pada data yang sama (Rousseeuw, 1987).

SC merupakan metode gabungan dari metode *cohessian* dan *separation*. Metode *cohessian* digunakan untuk menghitung seluruh data yang berada dalam kelompok yang sama. Sedangkan *separation* digunakan untuk menghitung jarak rata-rata tiap data dalam sebuah kelompok dengan kelompok lain. Untuk menghitung jarak antara data digunakan persamaan jarak *euclidean*. Nilai *Silhouette* berada pada rentang -1 hingga 1. Semakin besar nilai SC maka semakin baik kualitas pengelompokan yang dihasilkan (Kaufman & Rousseeuw, 1990). Langkah-langkah menghitung nilai SC yaitu sebagai berikut:

1. Menghitung jarak rata-rata data pada kelompok yang sama ($a(i)$). Misalkan data ke-i merupakan anggota pada kelompok ke-l maka $a(i)$ didefinisikan sebagai jarak rata-rata data ke-i dengan semua data yang berada di kelompok ke-l.

$$a(i) = \frac{1}{n_l - 1} \sum_{k_l=1}^{n_l} d(i, k_l) \tag{2.14}$$

dimana,

$a(i)$	: Jarak rata-rata data pada kelompok yang sama
n_l	: Banyak data pada kekompok ke-l
$d(i, k_l)$	: Jarak antara data ke-i dengan data ke-k pada kelompok l
k_l	: 1,2,3,...,n_l

2. Menghitung jarak rata-rata terkecil pada kelompok yang berbeda ($b(i)$). Misalkan data ke-i merupakan anggota pada kelompok ke-l maka $b(i)$ didefinisikan sebagai jarak rata-rata terkecil antara data ke-i dengan semua data pada kelompok yang berbeda (l').

$$b(i) = min \frac{1}{n_{l'}} \sum_{k_{l'}=1}^{n_{l'}} d(i, k_{l'}) \tag{2.15}$$

dimana,

$b(i)$	: Jarak rata-rata terkecil pada kelompok yang berbeda
$n_{l'}$	: Banyak data pada kelompok yang berbeda (l')
$d(i, k_{l'})$	: Jarak antara data ke-i dengan data ke-k pada kelompok l'

$k_{l'}$: 1,2,3,...,$n_{l'}$

3. Menghitung nilai *Silhouette* di tiap Kabupaten/Kota ($s(i)$)

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \tag{2.16}$$

4. Menghitung nilai SC

$$SC = \frac{1}{n}\sum_{i=1}^{n} s(i) \tag{2.17}$$

dimana,

n : Banyak sampel data

Sebuah tampilan grafis baru diusulkan untuk teknik partisi. Setiap kelompok diwakili oleh yang disebut *Silhouette*, yang didasarkan pada perbandingan keketatan dan pemisahannya. *Silhouette* ini menunjukkan data mana yang berada dengan baik di dalam kelompok mereka dan mana yang berada di antara kelompok. Seluruh pengelompokan ditampilkan dengan menggabungkan *Silhouette* ini dalam satu plot, memungkinkan penilaian terhadap kualitas relatif kelompok-kelompok dan gambaran umum dari konfigurasi data. Rata-rata *Silhouette* memberikan evaluasi validasi pengelompokan dan dapat digunakan untuk menentukan jumlah kelompok yang tepat (Rousseeuw, 1987).

2.5 *Stunting*

Masalah gizi dapat terjadi pada anak sejak berada di dalam kandungan hingga usia dua tahun. Pertumbuhan dan perkembangan yang sehat bergantung pada nutrisi yang cukup dan seimbang pada 1.000 hari pertama kehidupan. Oleh karena itu, status kesehatan dan gizi ibu merupakan faktor penting dari terjadinya *stunting* pada anak (Taki, 2018).

Stunting (pendek) merupakan masalah gizi akibat kekurangan asupan gizi yang berlangsung lama. Kejadian *stunting* pada anak merupakan masalah gizi utama yang dihadapi di Indonesia khususnya Provinsi Papua. Hal ini dapat dilihat dari hasil pengelompokan provinsi berdasarkan faktor penyebab balita *stunting* yang dilakukan oleh Satriawan dan Styawan (2021). Hasil tersebut menyatakan bahwa Provinsi Papua merupakan salah satu provinsi yang tergolong dalam

kelompok dengan faktor *stunting* sangat tinggi. Berdasarkan data SSGI, prevalensi *stunting* pada tahun 2020 sebesar 26,9% serta tahun 2021 dan 2022 masing-masing sebesar 24,4% dan 21,6% (Kementrian Kesehatan RI, 2023). Terdapat beberapa faktor yang dapat menjadi penyebab terjadinya *stunting* di Indonesia, diantaranya :

1. Pemberian ASI Eksklusif Kurang Dari 6 Bulan Pertama

   ASI mengandung zat gizi yang diperlukan oleh bayi untuk pertumbuhan dan perkembangan bayi (Hayati & Goejantoro, 2020). Selain itu, ASI dapat meningkatkan daya tahan tubuh, perlindungan penyakit infeksi dan perlindungan alergi sehingga apabila pemberian ASI Eksklusif kurang dari 6 bulan pertama maka akan meningkatkan risiko kejadian *stunting* (Sumarni, dkk, 2020).

2. Berat Badan Lahir Rendah (BBLR)

   Bayi dengan riwayat berat badan bayi saat lahir kurang dari 2500 gram atau Berat Badan Lahir Rendah (BBLR) akan memiliki resiko lebih tinggi mengalami keterlambatan perkembangan dan pertumbuhan karena bayi BBLR telah mengalami *Intrauterine Growth Restriction* yang menyebabkan pertumbuhan dan perkembangan lebih lambat (Kamilia, 2019).

3. Inisiasi Menyusui Dini (IMD)

   Penyebab masalah *stunting* salah satunya akibat dari penundaan Inisiasi Menyusui Dini (IMD). Anak yang tidak mendapatkan IMD memiliki kemungkinan 2,63 kali lebih tinggi mengalami kejadian *stunting* (Permadi, dkk, 2016).

4. Vitamin A

   Kekurangan vitamin A dapat menyebabkan terganggunya fungsi pertumbuhan sehingga tinggi balita lebih rendah dari tinggi anak seusianya. Kurangnya asupan vitamin A pada balita dapat menyebabkan resiko 17,5 kali terkena *stunting* dibandingkan balita yang mendapatkan asupan vitamin A yang cukup (Sari, dkk, 2022).

5. Keluarga Yang Tidak Memiliki Jaminan Pelayanan Kesehatan

   Kepemilikan jaminan pelayanan kesehatan merupakan salah satu faktor tidak langsung penyebab *stunting*. Balita yang memiliki jaminan pelayanan kesehatan memiliki kemudahan dalam mengakses pelayanan kesehatan sehingga dapat mengurangi resiko terjadinya *stunting* pada balita (Pertiwi, dkk, 2021).

6. Ibu Hamil Mengkonsumsi Tablet Tambah Darah < 90 tablet

   Studi yang menggunakan data dari hasil Riset Kesehatan Dasar (Riskesdas) tahun 2013, pemberian tablet tambah darah pada ibu hamil merupakan salah satu upaya percepatan pencegahan *stunting* di Indonesia. Studi menemukan bahwa ada hubungan konsumsi TTD pada ibu hamil ≥ 90 tablet dengan *stunting*. Ibu yang mengonsumsi TTD <90 tablet berpeluang 1,05 kali memiliki anak *stunting* dibanding ibu yang mengonsumsi TTD ≥ 90 tablet (Fentiana & Ginting, 2022).

7. Sumber Air Minum Tidak Layak

   Sumber air bersih yang tidak memadai merupakan salah satu faktor sanitasi lingkungan yang buruk. Kondisi sanitasi yang buruk dapat meningkatkan resiko terjadinya penyakit infeksi yang dapat mengganggu proses penyerapan nutrisi sehingga dapat menyebabkan *stunting* pada balita (Angraini, dkk, 2021).

2.6 Penelitian Terdahulu

Adapun penelitian relevan yang telah dilakukan sebelumnya dan menjadi rujukan dalam penelitian ini antara lain :

1. Hasil penelitian Sitohang, dkk (2019) dengan menggunakan metode *Fuzzy K-Means* menyatakan bahwa terdapat 514 Kabupaten/Kota di Indonesia masuk ke dalam kategori permasalahan kurang gizi (*underweight, wasting, stunting*) menunjukkan bahwa ada sebanyak 192 (37,4%) Kabupaten/Kota di Indonesia masuk ke dalam kategori rendah, 178 (28%) Kabupaten/Kota dalam kategori sedang, dan 144 (34,7%) Kabupaten/Kota dalam kategori tinggi. Sementara itu, hasil uji beda terhadap intervensi spesifik gizi di tiap klaster menggambarkan Inisiasi Menyusui Dini (IMD), pemberian Air Susu Ibu (ASI) eksklusif,

Pemberian Makanan Tambahan (PMT) pada balita kurus dan ibu hamil Kekurangan Energi Kronik (KEK), serta pemberian Tablet Tambah Darah (TTD) memiliki perbedaan yang signifikan secara statistik di setiap kelompok. Selanjutnya, intervensi sensitif gizi yang memiliki perbedaan signifikan di tiap kelompok adalah kepemilikan sanitasi layak oleh rumah tangga serta rumah tangga yang memiliki fasilitas mencuci tangan pakai sabun. Peningkatan implementasi intervensi permasalahan gizi spesifik dan sensitif perlu dilakukan di setiap daerah dengan memperhatikan karakteristik permasalahan gizi yang ada.

2. Hasil penelitian Satriawan dan Styawan (2021) dengan menggunakan metode Ward menyatakan bahwa terdapat empat kelompok dengan karakteristik yang berbeda-beda. Kelompok pertama beranggotakan 16 provinsi merupakan kelompok dengan faktor *stunting* tinggi. Kelompok kedua beranggotakan 8 provinsi merupakan kelompok dengan faktor *stunting* sedang. Kelompok ketiga beranggotakan 6 provinsi merupakan kelompok dengan faktor *stunting* rendah. Kelompok empat beranggotakan 4 provinsi merupakan kelompok dengan faktor *stunting* sangat tinggi.
3. Hasil penelitian Fadilah, dkk. (2022) dengan menggunakan metode *K-Means* dan dengan bantuan metode Elbow menghasilkan 2 *cluster* sebagai *cluster* terbaik dengan nilai selisih *Sum of Square Error* (SSE) sebesar 1401,51 dimana *cluster* 1 merupakan *cluster* dengan faktor penyebab *stunting* tinggi yang terdiri dari 324 Kabupaten/Kota, dan *cluster* 2 merupakan *cluster* dengan faktor penyebab *stunting* rendah yang terdiri dari 49 Kabupaten/Kota.
4. Hasil penelitian Rahma (2023) menunjukkan bahwa dari hasil uji *Silhouette Coefficient Index* diperoleh bahwa pada metode *Subtractive Fuzzy C-Means* jumlah klaster optimal yang terbentuk adalah dua kelompok yang terdapat pada jari-jari 0,85. Sedangkan pada metode *K-Means* jumlah klaster optimal yaitu dua kelompok. Berdasarkan uji *Silhouette Coefficient Index* juga diperoleh bahwa metode yang lebih baik pada penelitian ini adalah metode *K-Means* karena nilai *Silhouette Coefficient Index* yang diperoleh lebih tinggi dari metode *Subtractive Fuzzy C-Means*.