

BAB II

LANDASAN TEORI

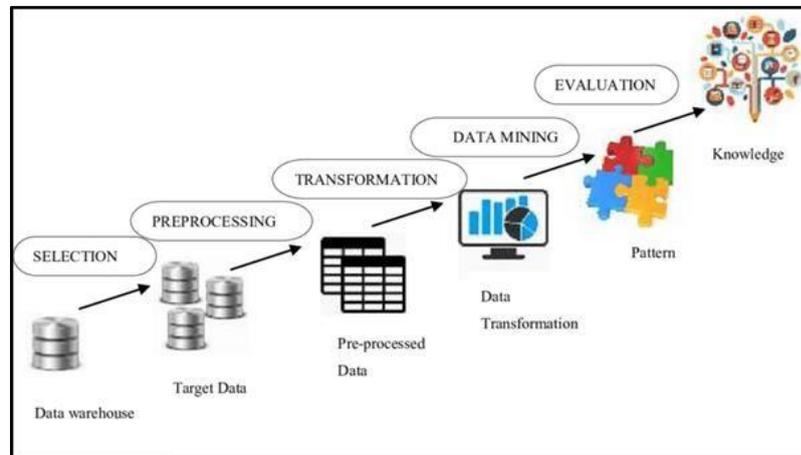
2.1 *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database adalah proses yang mencakup seleksi, pembersihan, transformasi, dan pengolahan data yang bertujuan untuk menemukan pola yang berguna. Mereka menekankan bahwa KDD adalah proses yang melibatkan eksplorasi data dalam konteks masalah spesifik untuk menghasilkan pengetahuan yang bermakna dan berguna dalam pengambilan Keputusan [1].

Tahapan Utama yang diambil untuk menemukan informasi yang bermanfaat dari data. Mereka melihat KDD sebagai rangkaian dari serangkaian proses, mulai dari pengumpulan data hingga penerapan teknik data mining untuk menggali pengetahuan yang terkandung dalam data tersebut. Dalam konteks ini, mereka juga menyebutkan pentingnya pengolahan data besar (Big Data) dalam KDD [2].

KDD sebagai proses yang menggabungkan data mining dan statistik untuk mengeksplorasi data secara lebih mendalam dan menemukan pengetahuan yang tidak dapat ditemukan hanya dengan metode tradisional. Dalam pandangannya, KDD adalah pendekatan yang lebih fleksibel yang dapat diterapkan dalam berbagai bidang, seperti kesehatan, bisnis, dan ilmu sosial[3].

Dalam konteks penelitian kesehatan, KDD memungkinkan analisis data yang kompleks menjadi lebih efektif. Proses iteratif KDD memberikan kesempatan untuk menyempurnakan model dan meningkatkan akurasi prediksi, yang dapat mendukung implementasi intervensi berbasis data[4].



Gambar 2.1 Diagram Tahapan Proses *Knowledge Discoveryin Database*

Sumber : Putri al. 2024

2.1.1 Tahapan Proses *Knowledge Discoveryin Database*

1. Seleksi Data

Tahapan seleksi data bertujuan untuk memilih data yang relevan dari basis data operasional. Proses ini melibatkan identifikasi atribut-atribut penting yang sesuai dengan tujuan penelitian, seperti variabel gizi, pendidikan, atau sanitasi dalam penelitian stunting. Data yang dipilih kemudian disimpan secara terpisah untuk dianalisis lebih lanjut[5]

2. Prapemrosesan Data (Preprocessing/Cleaning)

Prapemrosesan data adalah langkah krusial untuk membersihkan data dari kesalahan, duplikasi, dan inkonsistensi. Langkah ini juga mencakup proses enrichment, yaitu memperkaya data dengan informasi tambahan yang relevan dari sumber eksternal. Normalisasi dilakukan untuk menyelaraskan format data agar sesuai dengan kebutuhan analisis [6].

3. Transformasi Data

Transformasi data adalah proses mengubah data mentah menjadi format yang sesuai untuk proses data mining. Transformasi melibatkan pengkodean atribut, pembuatan variabel baru, dan pengelompokan data untuk memastikan data dapat dianalisis secara optimal[7].

4. Data Mining

Data mining merupakan inti dari KDD, di mana algoritma tertentu digunakan untuk mengekstrak pola atau informasi dari data yang telah diproses. Teknik atau algoritma yang digunakan dalam data mining sangat bergantung pada tujuan analisis. Dalam penelitian ini, algoritma apriori digunakan untuk mengklasifikasikan faktor risiko stunting berdasarkan variabel kesehatan, gizi, dan lingkungan[8].

5. Evaluasi dan Interpretasi Hasil

Evaluasi adalah tahapan akhir untuk memastikan bahwa pola yang ditemukan valid, relevan, dan dapat digunakan untuk pengambilan keputusan. Interpretasi dilakukan dengan menyajikan hasil dalam format yang mudah dipahami oleh pemangku kepentingan[9].

2.2 Data Mining

Data mining adalah proses inti dalam Knowledge Discovery in Databases (KDD) yang bertujuan untuk menggali informasi, membangun model, serta menemukan pola yang tersembunyi dalam kumpulan data yang besar. Proses ini seringkali melibatkan penerapan teknik matematika, statistik, dan teknologi untuk mengekstraksi data dan mengidentifikasi hubungan antar data yang bermanfaat.

Data mining memungkinkan pengolahan data dalam jumlah besar untuk menghasilkan informasi yang relevan dan dapat diterapkan pada pemecahan masalah tertentu[10]. Seiring berkembangnya teknologi, data mining menjadi lebih terintegrasi dengan sistem informasi yang ada, sehingga memudahkan dalam mengekstrak pola atau tren yang dapat digunakan untuk pengambilan keputusan. Oleh karena itu, Teknik ini sangat relevan dalam konteks penelitian kesehatan, di mana data yang kompleks dapat dianalisis untuk memprediksi risiko penyakit atau faktor lain yang berhubungan dengan kesehatan masyarakat[11].

2.2.1 Fungsi Data Mining

Data mining memiliki berbagai fungsi utama yang bermanfaat dalam menganalisis data, antara lain

1. **Klasifikasi**

Klasifikasi adalah teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan kategori tertentu. Dalam konteks penelitian kesehatan, klasifikasi dapat digunakan untuk memprediksi status kesehatan atau risiko penyakit, seperti stunting, berdasarkan variabel-variabel yang ada[12].

2. **Clustering (Pengelompokan)**

Clustering adalah proses pengelompokan data yang belum diberi label berdasarkan kesamaan karakteristik. Teknik ini digunakan untuk mengidentifikasi kelompok data yang memiliki kemiripan yang tinggi satu sama lain. Dalam penelitian ini, clustering dapat digunakan untuk

mengelompokkan individu berdasarkan faktor-faktor seperti status gizi dan kondisi sosial ekonomi mereka, yang dapat membantu dalam menganalisis revalensi stunting[13].

3. Asosiasi

Asosiasi adalah teknik untuk menemukan hubungan antara atribut dalam sebuah dataset. Sebagai contoh, asosiasi dapat digunakan untuk mengidentifikasi pola hubungan antara pola makan dengan tingkat gizi pada anak, yang berkontribusi pada terjadinya stunting[14].

2.2.2 Tahapan Data Mining

Data mining memiliki beberapa tahapan yang digunakan untuk menggali dan menganalisis data secara lebih mendalam. Berikut adalah tahapan-tahapan data mining yang relevan dalam penelitian ini

1. Deskripsi

Pada tahapan ini, peneliti berusaha menggambarkan pola atau kecenderungan yang ada dalam data tanpa membuat prediksi atau estimasi]. lebih lanjut. Ini adalah langkah awal untuk memahami data dan mengidentifikasi pola dasar[15].

2. Estimasi

Estimasi berfokus pada prediksi nilai variabel target berdasarkan nilai variabel prediktor. Dalam penelitian kesehatan, estimasi digunakan untuk memprediksi hasil berdasarkan variabel seperti usia, status gizi, dan pola makan.

3. Prediksi

Prediksi melibatkan penggunaan data untuk meramalkan hasil yang akan terjadi di masa depan. Misalnya, dalam konteks stunting, prediksi dapat dilakukan untuk mengidentifikasi kemungkinan seorang anak mengalami stunting berdasarkan faktor risiko yang ada[16].

4. Klasifikasi

Klasifikasi digunakan untuk mengelompokkan data ke dalam kategori yang sudah ditentukan sebelumnya. Sebagai contoh, dalam penelitian ini, klasifikasi dapat digunakan untuk mengelompokkan anak-anak berdasarkan tingkat risiko stunting menjadi kategoritinggi, sedang, dan rendah. Proses ini mempermudah pengambilan keputusan untuk tindakan lebih lanjut[17].

5. Pengelompokan

Pengelompokan adalah teknik untuk mengelompokkan data tanpa menggunakan variabel target. Misalnya, data tentang anak-anak dapat dikelompokkan berdasarkan faktor-faktor seperti pola makan, status gizi, atau tingkat sanitasi yang mereka alami. Pengelompokan ini memungkinkan analisis lebih dalam tentang prevalensi stunting di kelompok tertentu[18].

6. Asosiasi

Pengelompokan adalah teknik untuk mengelompokkan data tanpa menggunakan variabel target. Misalnya, data tentang anak-anak dapat dikelompokkan berdasarkan faktor faktor seperti pola makan, status gizi, atau tingkat sanitasi yang mereka alami. Pengelompokan ini

memungkinkan analisis lebih dalam tentang prevalensi stunting di kelompok tertentu[19].

2.3 Stunting

Stunting adalah kondisi gagal tumbuh pada anak usia dini yang ditandai dengan tinggi badan yang lebih rendah dari standar umur yang sesuai. Stunting sering kali disebabkan oleh kurangnya asupan gizi yang tepat pada usia 1000 Hari Pertama Kehidupan (HPK), yang mencakup periode kehamilan hingga dua tahun pertama kehidupan. Kondisi ini tidak hanya berdampak pada pertumbuhan fisik anak, tetapi juga dapat memengaruhi perkembangan kognitif dan sosial anak, yang pada akhirnya dapat mempengaruhi kualitas hidup mereka dimasa depan[20]. Stunting merupakan gambaran dari status gizi kurang pada anak balita yang bersifat kronik. Kondisi ini terjadi selama masa pertumbuhan dan perkembangan anak, sehingga menyebabkan anak menjadi lebih pendek dibandingkan dengan anak seusianya. Penentuan status stunting dapat dilakukan dengan menggunakan perhitungan Z-score untuk mengukur tinggi badan anak. Z-score membandingkan tinggi badan anak dengan standar referensi yang ada, dan hasilnya dikategorikan dalam dua kategori:

1. Kategori Pendek: Z-score antara -3 SD hingga $< -2SD$
2. Kategori Sangat Pendek : Z-score $< -3SD$

Stunting yang terjadi pada anak usia dini dapat berisiko terhadap kualitas hidup mereka dalam jangka panjang. Anak-anak yang mengalami stunting berisiko lebih tinggi untuk mengalami keterlambatan dalam perkembangan fisik dan kognitif, yang pada gilirannya dapat mempengaruhi prestasi akademik mereka dan

kualitas hidup di masa depan[21].

2.3.1 Faktor-faktor Penyebab Stunting

Salah satu faktor utama yang mempengaruhi stunting adalah status sosial ekonomi[22]. Menemukan bahwa anak-anak yang berasal dari keluarga berpendapatan rendah memiliki risiko yang lebih tinggi untuk mengalami stunting, dimana keluarga yang berada pada strata ekonomi terendah memiliki kemungkinan 5,39 kali lebih besar untuk anak anaknya mengalami stunting dibandingkan dengan keluarga yang berasal dari latar belakang ekonomi yang lebih tinggi. Hal ini dilakukan oleh [23], yang melaporkan bahwa anak anak dari keluarga berpendapatan rendah memiliki risiko 43,1 kali lebih tinggi untuk mengalami stunting dibandingkan anak-anak dari keluarga berpendapatan menengah. Hubungan antara kemiskinan dan stunting sudah terdokumentasi dengan baik, karena keluarga berpenghasilan rendah seringkali memiliki akses terbatas terhadap makanan bergizi, sehingga menyebabkan asupan makanan tidak memadai [24].

Penentu penting lainnya adalah pendidikan dan kesehatan ibu. Penelitian menunjukkan bahwa rendahnya tingkat pendidikan ibu berkorelasi dengan tingginya angka stunting pada anak. [25] menggaris bawahi bahwa ibu dengan latar belakang pendidikan terbatas cenderung tidak melakukan praktik pemberian ASI eksklusif, yang penting untuk pertumbuhan anak yang optimal. Lebih lanjut [26] menunjukkan bahwa tinggi badan ibu juga merupakan faktor yang signifikan.

Ibu yang memiliki tinggi badan kurang dari 150 cm mempunyai peluang 2,5

kali lebih tinggi untuk mempunyai anak stunting dibandingkan dengan ibu yang tinggi badannya. Selain faktor sosio-ekonomi dan ibu, masalah kesehatan seperti infeksi dan penyakit merupakan kontributor signifikan terhadap stunting. Menurut [27], penyakit yang sering diderita, termasuk diare dan infeksi saluran pernafasan, dapat mengganggu status gizi dan berkontribusi terhadap stunting.

Hubungan dua arah antara gizi buruk dan infeksi berarti bahwa anak-anak yang kekurangan gizi lebih rentan terhadap infeksi, yang selanjutnya memperburuk kekurangan gizi mereka. [28] juga mencatat bahwa prevalensi stunting tertinggi terjadi pada anak berusia 24-59 bulan, periode ketika mereka sangat rentan terhadap infeksi akibat pengenalan makanan pendamping ASI dan penurunan pemberian ASI.

1. Gizi Ibu

Status gizi ibu selama kehamilan berpengaruh langsung terhadap pertumbuhan janin. Ibu dengan status gizi yang kurang atau kurangnya asupan nutrisi dapat melahirkan bayi dengan berat badan lahir rendah (BBLR), yang berisiko tinggi mengalami stunting di masa depan.

2. Penyakit Infeksi

Penyakit infeksi, seperti diare dan infeksi saluran pernapasan, dapat mengganggu penyerapan nutrisi pada anak, yang mempengaruhi pertumbuhannya. Penyakit ini mengurangi ketersediaan gizi yang diperlukan untuk perkembangan tubuh anak.

3. Faktor Sosial Ekonomi

Pendidikan ibu dan kondisi ekonomi keluarga berperan penting dalam

menentukan status gizi anak. Ibu dengan tingkat pendidikan yang lebih rendah mungkin memiliki pengetahuan yang terbatas mengenai pentingnya gizi yang seimbang dan perawatan kesehatan anak. Kondisi ekonomi keluarga yang kurang stabil sering menghambat akses terhadap makanan bergizi dan layanan kesehatan yang diperlukan.

4. Lingkungan dan Sanitasi

Lingkungan yang tidak mendukung, seperti akses terbatas terhadap air bersih dan sanitasi yang buruk, dapat meningkatkan risiko infeksi yang mempengaruhi status gizi anak. Kesehatan lingkungan yang buruk dapat memperburuk kondisi stunting pada anak-anak, karena infeksi yang sering terjadi dapat mengurangi kemampuan tubuh untuk menyerap nutrisi secara optimal[29].

2.3.2 Intervensi untuk Mencegah dan Mengatasi Stunting

Untuk mengatasi masalah stunting, dibutuhkan pendekatan yang komprehensif yang melibatkan intervensi diberbagai sektor. Intervensi ini tidak hanya berfokus pada perbaikan gizi, tetapi juga mencakup aspek kesehatan, pendidikan, dan lingkungan. Berikut adalah beberapa langkah intervensi yang perlu diambil

1. Pendidikan Gizi dan Pemberdayaan Ibu

Penyuluhan tentang pentingnya gizi selama kehamilan, menyusui, dan pemberian makanan pendamping ASI (MP-ASI) yang sesuai sangat penting untuk mencegah stunting. Pendidikan tentang cara memenuhi kebutuhan gizi yang tepat dapat membantu ibu memberikan asupan

yang cukup untuk pertumbuhan optimal anak.

2. Peningkatan Akses Kesehatan

Akses terhadap layanan kesehatan yang memadai harus diperbaiki, terutama untuk deteksi dini dan penanganan penyakit infeksi yang dapat memengaruhi status gizi anak. Program imunisasi dan pengobatan terhadap infeksi juga sangat penting untuk mencegah infeksi yang dapat memperburuk status gizi anak.

3. Perbaiki Sanitasi dan Akses Air Bersih

Pemerintah dan Masyarakat harus bekerjasama untuk meningkatkan sanitasi lingkungan dan akses terhadap air bersih. Lingkungan yang sehat berperan penting dalam pencegahan penyakit yang dapat mengganggu pertumbuhan anak.

4. Pendekatan Multi sektoral

Stunting adalah masalah yang kompleks dan membutuhkan pendekatan multi sektoral yang melibatkan sektor kesehatan, pendidikan, dan ekonomi. Program-program yang terintegrasi, yang mengatasi masalah gizi, akses kesehatan, dan lingkungan hidup, akan lebih efektif dalam mengurangi prevalensi stunting di Indonesia

2.4 Algoritma Apriori

Algoritma Apriori adalah algoritma pencarian pola asosiasi yang digunakan untuk menemukan itemset yang sering muncul dalam dataset besar. Algoritma ini beroperasi berdasarkan prinsip apriori property, yang menyatakan bahwa jika suatu itemset sering muncul dalam data, maka semua subset dari itemset tersebut

juga harus sering muncul. Apriori digunakan untuk mengekstraksi hubungan atau pola yang tersembunyi, yang sering diterapkan dalam analisis keranjang belanja di e-commerce atau analisis lainnya[30]. Apriori adalah algoritma iteratif yang mengidentifikasi itemset yang sering dengan menggunakan strategi bottom-up, yang berarti itemset kecil terlebih dahulu diperiksa sebelum membentuk itemset yang lebih besar. Mereka menekankan bahwa meskipun Apriori sederhana, algoritma ini dapat menghadapi tantangan terkait kompleksitas komputasi pada data besar, terutama ketika jumlah kandidat itemset sangat besar[31]. Algoritma Apriori adalah algoritma yang menggunakan pendekatan brute-force untuk mencari asosiasi antara item-item dalam data. Mereka menyoroti bahwa dengan pruning atau pemangkasan itemset yang tidak memenuhi kriteria, algoritma ini dapat mengurangi jumlah kandidat yang perlu dievaluasi. Namun, mereka juga menunjukkan bahwa algoritma Apriori lebih cocok untuk dataset kecil hingga sedang, sementara pada dataset yang lebih besar diperlukan teknik tambahan untuk meningkatkan efisiensi[32].

Rumus Algoritma Apriori:

$$Support(X) = \frac{\text{Jumlah transaksi yang mengandung } X}{\text{Total jumlah transaksi}}$$

Gambar 2.2 Rumus Support

Sumber : Amin et al, 2021 Dimana:

a. Support

Proporsi transaksi yang mengandung item X, menunjukkan seberapa sering item tersebut muncul dalam keseluruhan dataset.

- b. Jumlah Transaksi yang mengandung X

Banyaknya transaksi dalam dataset yang mencakup item X.

- c. Total Jumlah Transaksi

Keseluruhan transaksi dalam dataset yang sedang dianalisis

$$\text{Confidence}(X \Rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

Gambar 2.3 Rumus Confidence

Sumber : Amin et al, 2021

Dimana

- a. Confidence ($X \Rightarrow Y$)

Probabilitas bahwa jika X terjadi, maka Y juga akan terjadi dalam satu transaksi.

- b. Support ($X \cup Y$)

Persentase transaksi yang mengandung X dan Y secara bersamaan.

- c. Support (X)

Persentase transaksi yang mengandung X.

2.5 Naive Bayes

Naive Bayes adalah algoritma klasifikasi yang berbasis pada teori probabilitas Bayes dengan asumsi bahwa fitur-fitur yang digunakan dalam model adalah independen satu sama lain, yang dikenal dengan nama "naive" (naif). Algoritma ini bekerja dengan cara menghitung probabilitas posterior suatu kelas berdasarkan data yang tersedia, lalu memilih kelas dengan probabilitas tertinggi. Walaupun asumsi independensi fitur ini jarang berlaku dalam kenyataan, Naive Bayes tetap

memberikan hasil yang baik dalam banyak kasus, terutama untuk klasifikasi teks[33]. Naive Bayes adalah model klasifikasi yang sangat populer karena kesederhanaan dan efisiensinya dalam klasifikasi data yang tidak terlalu kompleks. Algoritma ini mengandalkan asumsi independensi antar fitur, yang membuatnya lebih mudah diterapkan dibandingkan dengan model klasifikasi lainnya seperti SVM (Support Vector Machine) atau decision trees[34].

Rumus Naive Bayes

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

di mana:

- $P(C|X)$ = Probabilitas bahwa suatu data termasuk dalam kelas C, diberikan fitur X (probabilitas posterior).
- $P(X|C)$ = Probabilitas bahwa fitur X muncul jika kelasnya C (likelihood).
- $P(C)$ = Probabilitas kelas C dalam dataset (prior probability).
- $P(X)$ = Probabilitas munculnya fitur X (evidence).

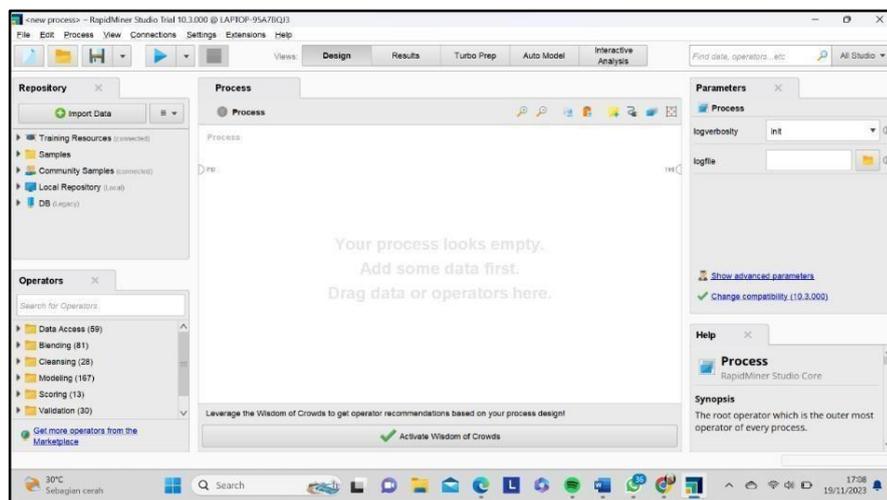
Gambar 2.4 Rumus Naive Bayes

Sumber : Gambhir & Kaur (2021)

2.6 Rapid Miner

Rapid Miner adalah alat perangkat lunak yang menyediakan platform integrasi data, pemrosesan data, pembelajaran mesin, dan analisis statistik dalam satu lingkungan kerja yang terintegrasi. Salah satu fitur utama dari Rapid Miner adalah kemampuannya untuk memungkinkan pengguna, baik yang berpengalaman maupun yang baru dalam data science, untuk melakukan eksperimen dan pengolahan data tanpa memerlukan keterampilan pemrograman yang mendalam[35][36] Menggambarkan Rapid Miner sebagai platform open-

source yang menyediakan berbagai alat untuk data mining dan machine learning. Mereka menyoroti bahwa RapidMiner mendukung berbagai teknik analisis, seperti klasifikasi, regresi, klustering, dan analisis teks. Platform ini dapat menghubungkan berbagai sumber data, termasuk database, file CSV, atau data real-time, serta memungkinkan integrasi dengan berbagai bahasa pemrograman, seperti Python dan R, untuk memberikan fleksibilitas dalam analisis.



Gambar 2.5 Tampilan utama Rapid Miner

Sumber : Lilly. Staff

2.7 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

Referensi Penelitian	1
Judul	Strategi Pemerintah Desa dalam Penanganan Stunting di Desa Panduman, Kecamatan Jelbuk, Kabupaten Jember

Nama Penulis	Muhammad Syukron Abdillah,Ria Angin dan Putri Robiatul Adawiyah
Tahun	2024
Hasil	Stunting merupakan kondisi gagal pertumbuhan pada anak balita akibat kekurangan gizi kronis, yang dapat menyebabkan dampak kesehatan, sosial, dan ekonomi di masa depan.
Referensi Penelitian	2
Judul	Gerakan Pencegahan Stunting Melalui Edukasi pada Masyarakat di Desa Muntoi Kabupaten Bolaang Mongondow
Nama Penulis	St. Rahmawati Hamzah , Hamzah
Tahun	2020
Hasil	Kasus stunting di Indonesia masih tergolong tinggi dengan menyentuh angka 30,8%. Angka ini masih jauh dari standar WHO yaitu 20%.
Referensi Penelitian	3
Judul	PENCEGAHAN STUNTING MELALUI EDUKASI MAKANAN BERGIZI & BERAGAM DI DESA KARANGSIDEMEN

Nama Penulis	Hayati,Achmad Wisnu Adi Riyanto, Zamrutin Fata, Ahmad Doni
Tahun	2023
Hasil	Kegiatan verifikasi data kasus stunting telah dilakukan di 14 posyandu yang terdapat di setiap dusun di Desa Karang Sidemen.
Referensi Penelitian	4
Judul	Peran Kader dalam Penurunan Stunting di Desa
Nama Penulis	KadarRamadhan,ChristinaEntoh, Nurfatimah
Tahun	2022
Hasil	Keseriusan pemerintah dalam pencegahan stunting yang terdapat dalam strategi nasional percepatan pencegahan stunting Patut diapresiasi.
Referensi Penelitian	5
Judul	Hubungan karakteristik keluarga dengan kejadian stunting pada anak 6 – 24 bulan di Kabupaten Langka
Nama Penulis	Sri Henny Siregar , Albiner Siagian
Tahun	2021

Hasil	Stunting merupakan masalah malnutrisi kronis akibat kekurangan gizi jangka panjang terkait dengan kurangnya asupan makanan.
-------	-----------------------------------------------------------------------------------------------------------------------------