

**KLASIFIKASI RISIKO GIZI BURUK PADA BALITA
MENGUNAKAN METODE C4.5 (STUDI KASUS:
UPTD PUSKESMAS GUNUNG MERIAH)**

SKRIPSI

DISUSUN OLEH

FATHUR RIZQI HASUGIAN

NPM. 2209010209



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2026**

**KLASIFIKASI RISIKO GIZI BURUK PADA BALITA
MENGUNAKAN METODE C4.5 (STUDI KASUS:
UPTD PUSKESMAS GUNUNG MERIAH)**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

**FATHUR RIZQI HASUGIAN
NPM. 2209010209**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2026**

LEMBAR PENGESAHAN

Judul Skripsi : KLASIFIKASI RISIKO GIZI BURUK PADA
- BALITA MENGGUNAKAN METODE C4.5
(STUDI KASUS: UPTD PUSKESMAS GUNUNG
MERIAH)
Nama Mahasiswa : FATHUR RIZQI HASUGIAN
NPM : 2209010209
Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



Dr. Firahmi Rizky, S.Kom., M.Kom
NIDN. 0116079201

Ketua Program Studi



Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom.
NIDN. 0118088902



Dekan



Prof. Dr. Al-Khowarizmi, M.Kom.
NIDN. 0117099201

PERNYATAAN ORISINALITAS

KLASIFIKASI RISIKO GIZI BURUK PADA BALITA MENGUNAKAN METODE C4.5 (STUDI KASUS: UPTD PUSKESMAS GUNUNG MERIAH)

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 21 Agustus 2026

Yang membuat pernyataan



Fathur Rizqi Hasugian
NPM. 2209010209

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademik Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan di bawah ini:

Nama : Fathur Rizqi Hasugian

NPM : 2209010209

Program Studi : Sistem Informasi

Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**KLASIFIKASI RISIKO GIZI BURUK PADA BALITA
MENGUNAKAN METODE C4.5 (STUDI KASUS:
UPTD PUSKESMAS GUNUNG MERIAH)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk *database*, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 21 Agustus 2026

Yang membuat pernyataan



Fathur Rizqi Hasugian
NPM. 2209010209

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Fathur Rizqi Hasugian
Tempat dan Tanggal Lahir : Tulaan, 27 Juni 2004
Alamat Rumah : Jln. Perjuangan
Telepon/Faks/HP : 082210851883
E-mail : rizqifathur56@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SDN TULAAAN TAMAT : 2016
SMP : MTSS MUHAMMADIYAH TAMAT: 2019
GUNUNG MERIAH
SMA : SMAS GUNUNG MERIAH TAMAT: 2022

KATA PENGANTAR



Assalamu 'alaikum warahmatullahi wabarakatuh

Alhamdulillahirabbil'alamin. Dengan nama Allah Yang Maha Pengasih lagi Maha Penyayang, puji syukur penulis panjatkan kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “SISTEM PENDUKUNG KEPUTUSAN BERBASIS GIS UNTUK OPTIMALISASI PEMANFAATAN RUANG USAHA MENGGUNAKAN METODE AHP ” dengan baik sebagai salah satu persyaratan untuk menyelesaikan jenjang strata satu (S1) dan untuk meraih gelar akademik Sarjana Komputer (S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Penulis menyadari bahwa skripsi ini tidak lepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak yang secara langsung maupun tidak langsung. Dengan segala kerendahan hati, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Akrim, M.Pd., selaku Rektor Universitas Muhammadiyah Sumatera Utara.
2. Bapak Assoc. Prof. Dr. Alkhowarizmi, M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi.
3. Ibu Dr. Firahmi Rizky, M.Kom., selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi.
4. Bapak Mhd. Basri, S.Si, M.Kom., selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.

5. Bapak Mahardika Abdi Prawira Tanjung, M.Kom., selaku Ketua Prodi Sistem Informasi.
6. Bapak Mulkan Azhari, S.Kom, M.Kom., selaku Sekretaris Prodi Sistem Informasi sekaligus dosen pembimbing akademik penulis selama perkuliahan.
7. Ibu Dr. Firahmi Rizky, M.Kom., selaku dosen pembimbing yang telah berkontribusi membantu, memberikan ide, saran, kritik dan bimbingan dalam menyelesaikan skripsi penulis
8. Ibu Bdn,Nurmila,S.Keb serta jajaran staf instansi puskesmas gunung meriah yang telah memberikan ruang untuk penulis belajar selama penelitian juga memberikan izin untuk meneliti dan mempermudah segala urusan penulis selama penelitian.
9. Teristimewa untuk kedua orang tua tersayang, ayah dan bunda, yang tiada hentinya melangitkan doa baiknya serta memberikan dukungan kepada penulis dalam memperjuangkan masa depan. Segala bentuk pengorbanan, kesabaran, dan cinta yang diberikan menjadi motivasi yang sangat berarti bagi penulis dalam menghadapi setiap proses, tantangan, dan kesulitan selama proses penyusunan skripsi ini hingga akhirnya dapat diselesaikan.
10. Teman-teman E1 Sistem Informasi 2022.
11. Teman-teman KKN 2025 Desa Paluh Kemiri, terima kasih atas segala dukungan, kebersamaan, cerita dan kenangan yang telah dilalui bersama. Terima kasih telah menjadi bagian dari perjalanan yang penuh makna dan semoga silaturahmi tetap terjaga. Untuk

seorang sahabat yang telah lebih dahulu menghadap Sang Pencipta,
semoga Allah SWT menerima seluruh amal ibadah dan
ditempatkan di tempat terbaik-Nya

Skripsi ini tentunya masih jauh dari kesempurnaan, untuk itu penulis
berharap kritik dan saran yang membangun untuk bahan pembelajaran
berkesinambungan di masa depan. Semoga Skripsi ini dapat bermanfaat
bagi penulis dan pembaca.

Wassalamu'alaikum warahmatullahi wabarakatuh.

DAFTAR ISI

DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	5
BAB II LANDASAN TEORI	7
2.1. Gizi Bayi dan Risiko Gizi Buruk	7
2.2. Konsep Dasar Decision Tree	8
2.3. Algoritma C4.5	10
2.3.1. Entropy.....	11
2.3.2. Mencari Nilai Gain	12
2.4. Perangkat Lunak (<i>Software</i>).....	12
2.4.1. Python	12
2.4.2. Python Flask	13
2.4.3. Scikit-Learn	14
2.4.4. Bootstrap.....	14
2.5. Unified Modeling Language (UML).....	15
2.5.1. Use Case Diagram	16
2.5.2. Activity Diagram	17
2.6. Literatur Review	17
BAB III ANALISIS DAN PERANCANGAN SISTEM	20
3.1. Analisis Permasalahan.....	20
3.2. Algoritma Sistem.....	22
3.2.1 Sumber Data	22
3.3. Tahapan Penelitian	24
3.3.1. Pengumpulan Data.....	26
3.3.2. Praproses Data	26
3.3.3. Penentuan Atribut dan Target	27

3.3.4. Penerapan Algoritma C4.5.....	27
3.3.5. Implementasi Sistem.....	28
3.3.6. Pengujian Sistem.....	28
3.3.7. Evaluasi dan Analisis Hasil	28
3.3.8. Penarikan Kesimpulan	29
3.4. Perhitungan Manual Algoritma C4.5	29
3.4.1. Distribusi Kelas Gizi.....	29
3.4.2. Perhitungan Atribut Umur	31
3.4.3. Perhitungan Atribut Jenis Kelamin.....	34
3.4.4. Perhitungan Atribut Berat Badan.....	37
3.4.5. Perhitungan Atribut Tinggi Badan.....	39
3.4.6. Perhitungan Atribut LILA	42
3.4.7. Perbandingan Nilai Gain Ratio	44
3.5. Perancangan Sistem.....	44
3.5.1. Use Case Diagram	45
3.5.2. Activity Diagram	46
3.5.3. Sequence Diagram	47
3.5.4. Class Diagram.....	49
3.5.5. Rancangan Antarmuka Aplikasi	53
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM	57
4.1. Implementasi Model dan Dataset	57
4.1.1. Implementasi Dataset Properti	57
4.2. Implementasi Prosesing Data	59
4.3. Implementasi Algoritma Decision Tree C4.5.....	60
4.4. Implementasi Antarmuka Sistem	62
4.4.1. Antarmuka Sistem Admin	62
4.4.2. Antarmuka Sistem User	67
4.5. Pengujian Sistem	70
4.5.1. Pengujian Sistem Menggunakan Black Box Testing.....	71
4.5.2 Pengujian Hasil Klasifikasi Menggunakan Confusion Matrix	73
BAB V KESIMPULAN DAN SARAN	75
5.1. Kesimpulan.....	75
5.2. Saran	76
DAFTAR PUSTAKA	77

DAFTAR TABEL

Tabel 2.6. Literatur Review	17
Tabel 3.1. Sumber Data.....	23
Tabel 3.2. Distribusi Kelas Status Gizi	29
Tabel 3.3. Kandidat Atribut Umur	31
Tabel 3.4. Distribusi Data Berdasarkan Umur	32
Tabel 3.5. Distribusi Data Berdasarkan Jenis Kelamin	36
Tabel 3.6. Distribusi Data Berdasarkan Berat Badan	37
Tabel 3.7. Kandidat Threshold Atribut Tinggi Badan	39
Tabel 3.8. Distribusi Berdasarkan Atribut Tinggi Badan	40
Tabel 3.9. Distribusi Data Berdasarkan atribut LILA.....	42
Tabel 3.10. Perbandingan Nilai Gain Ratio	44
Tabel 4.1. Pengujian BlackBox.....	71

DAFTAR GAMBAR

Gambar 3.1. Flowchart Tahapan Penelitian	25
Gambar 3.2. Use Case Diagram	46
Gambar 3.3. Activity Diagram	47
Gambar 3.4. Sequence Diagram	48
Gambar 3.5. Class Diagram	50
Gambar 3.6. Halaman Login	54
Gambar 3.7. Halaman Dashboard	54
Gambar 3.8. Halaman Data Balita	55
Gambar 3.9. Halaman Klasifikasi	56
Gambar 4.1. Dataset Balita	57
Gambar 4.2. Dataset Balita Pada MySQL	58
Gambar 4.3. implementasi Baca dataset dari MySQL	59
Gambar 4.4. Implementasi Preprocessing Data	59
Gambar 4.5. implementasi Model Decesion Three C4.5	60
Gambar 4.6. Rule Decesion Tree	61
Gambar 4.7. Halaman Login Admin	63
Gambar 4.8. Halaman Dashboard Admin	63
Gambar 4.9. Halaman Data Balita Admin	64
Gambar 4.10. Halaman Analisis Gizi Admin	65
Gambar 4.11. Halaman Hasil Analisis Admin	66
Gambar 4.12. Hasil Laporan PDF Admin	67
Gambar 4.13. Halaman Login User	67
Gambar 4.14. Halaman Dashboard User	68
Gambar 4.15. Halaman Analisis Gizi User	68
Gambar 4.16. Halaman Hasil Analisis Gizi	69
Gambar 4.17. Hasil Laporan PDF	70
Gambar 4.18. Script Pengujian Akurasi dan Confusion Matrix	73
Gambar 4.19. Confusion Matrix dan Accuracy	74

KLASIFIKASI RISIKO GIZI BURUK PADA BALITA MENGUNAKAN METODE C4.5 (STUDI KASUS: UPTD PUSKESMAS GUNUNG MERIAH)

ABSTRAK

Permasalahan gizi buruk pada balita merupakan tantangan kesehatan masyarakat yang signifikan, karena kekurangan gizi yang tidak terdeteksi sejak dini dapat menghambat pertumbuhan fisik dan perkembangan kognitif anak secara permanen. Di UPTD Puskesmas Gunung Meriah, proses identifikasi risiko gizi buruk masih mengandalkan pencatatan manual dan evaluasi parameter secara tradisional, yang cenderung memakan waktu serta rentan terhadap ketidakkonsistenan hasil penilaian. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi risiko gizi buruk menggunakan teknik *data mining* guna mengoptimalkan pemanfaatan data historis balita. Fokus utama penelitian adalah menerapkan algoritma C4.5 untuk membangun model klasifikasi, mengukur tingkat akurasi melalui *confusion matrix*, dan mengidentifikasi atribut antropometri yang paling dominan dalam menentukan status gizi. Metode penelitian mencakup pengumpulan data balita periode Oktober 2025 hingga Januari 2026, yang meliputi atribut umur, jenis kelamin, berat badan, tinggi badan, dan Lingkar Lengan Atas (LILA). Data tersebut diolah menggunakan algoritma C4.5 untuk menghasilkan pohon keputusan dan aturan klasifikasi yang mudah dipahami. Hasil penelitian menunjukkan bahwa model C4.5 berhasil mencapai tingkat akurasi sebesar 87,18% dalam memprediksi risiko gizi buruk. Atribut LILA teridentifikasi sebagai faktor yang paling berpengaruh terhadap klasifikasi dengan nilai *gain ratio* tertinggi sebesar 0,78623. Kesimpulan dari penelitian ini adalah algoritma C4.5 terbukti efektif dan objektif dalam mengklasifikasikan risiko gizi buruk pada balita. Implementasi sistem berbasis web ini memberikan kemudahan bagi tenaga kesehatan dalam melakukan deteksi dini dan menentukan prioritas penanganan, sehingga upaya pencegahan gizi buruk dapat dilakukan dengan lebih cepat dan tepat sasaran.

Kata Kunci: Gizi Buruk, Balita, Algoritma C4.5, *Data Mining*, Klasifikasi.

**CLASSIFICATION OF MALNUTRITION RISK IN
TODDLERS USING THE C4.5 METHOD (CASE STUDY:
UPTD HEALTH CENTER GUNUNG MERIAH**

ABSTRACT

Malnutrition in toddlers is a significant public health challenge, as undetected malnutrition can permanently hinder a child's physical growth and cognitive development. At the Gunung Meriah Community Health Center (UPTD), the process of identifying malnutrition risks still relies on manual recording and traditional parameter evaluation, which tends to be time-consuming and prone to inconsistencies in assessment results. This study aims to develop a malnutrition risk classification system using data mining techniques to optimize the use of historical toddler data. The main focus of the study is to apply the C4.5 algorithm to build a classification model, measure the level of accuracy through a confusion matrix, and identify the most dominant anthropometric attributes in determining nutritional status. The research method includes collecting toddler data from October 2025 to January 2026, which includes attributes of age, sex, weight, height, and Mid-Upper Arm Circumference (MUAC). The data is processed using the C4.5 algorithm to produce a decision tree and easy-to-understand classification rules. The results show that the C4.5 model successfully achieved an accuracy level of 87.18% in predicting the risk of malnutrition. The LILA attribute was identified as the most influential factor in classification, with the highest gain ratio value of 0.78623. The conclusion of this study is that the C4.5 algorithm is proven to be effective and objective in classifying the risk of malnutrition in toddlers. The implementation of this web-based system makes it easier for health workers to conduct early detection and determine treatment priorities, so that malnutrition prevention efforts can be carried out more quickly and precisely.

Keywords: Malnutrition, Toddlers, C4.5 Algorithm, Data Mining, Classification.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Permasalahan gizi buruk pada balita masih menjadi salah satu isu kesehatan masyarakat yang memerlukan perhatian serius, khususnya di negara berkembang. Status gizi balita merupakan indikator penting dalam menilai kualitas kesehatan dan pertumbuhan anak pada masa awal kehidupan. Kekurangan gizi yang tidak terdeteksi secara dini dapat berdampak pada gangguan pertumbuhan fisik, perkembangan kognitif, serta meningkatnya risiko penyakit pada usia dewasa. Oleh karena itu, deteksi dan klasifikasi risiko gizi buruk secara cepat dan tepat menjadi langkah strategis dalam upaya pencegahan dan penanganan masalah gizi.

Status gizi balita umumnya dinilai melalui pendekatan antropometri, seperti pengukuran berat badan, tinggi badan, dan umur. Hasil pengukuran tersebut kemudian diklasifikasikan ke dalam kategori tertentu seperti gizi buruk, gizi kurang, gizi normal, dan gizi lebih (Nuraisana et al., 2024). Klasifikasi ini menjadi dasar bagi tenaga kesehatan dalam menentukan langkah intervensi yang tepat, terutama dalam upaya pencegahan gizi buruk dan *stunting*.

UPTD Puskesmas Gunung Meriah merupakan salah satu fasilitas pelayanan kesehatan tingkat pertama yang memiliki peran penting dalam pemantauan pertumbuhan balita di wilayah kerjanya. Setiap periode pelayanan, puskesmas ini mengelola data balita yang meliputi umur, berat badan, tinggi badan, serta hasil penilaian status gizi. Data tersebut terus bertambah seiring meningkatnya jumlah balita yang melakukan pemeriksaan rutin.

Namun pada implementasinya, cara mengidentifikasi balita yang berisiko gizi buruk masih dilakukan dengan cara tradisional, yaitu mencatat secara manual dan mengevaluasi parameter tertentu satu per satu. Pendekatan ini biasanya memakan waktu lebih lama dan bisa menghasilkan penilaian yang tidak konsisten ketika jumlah data semakin banyak. Selain itu, kurangnya sistem yang mampu mengelola data historis balita secara terorganisir menyebabkan pola-pola penting yang tersembunyi dalam data belum dimanfaatkan secara maksimal sebagai dasar dalam pengambilan keputusan.

Seiring perkembangan teknologi informasi, pemanfaatan teknik *data mining* dapat menjadi solusi dalam mengolah data kesehatan secara sistematis. Menurut (Turban et al., 2011) *Data mining* adalah suatu proses pengolahan data dalam jumlah besar untuk menemukan pola serta informasi yang memiliki nilai guna. Informasi yang diperoleh dapat digunakan sebagai dasar dalam menentukan keputusan secara lebih tepat dan efisien. Salah satu teknik yang digunakan dalam data mining adalah klasifikasi, yaitu metode untuk mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristik atau atribut yang dimilikinya.

Konsep induksi pohon keputusan diperkenalkan oleh J. R. Quinlan melalui pendekatan *Top-Down Induction of Decision Trees* (TDIDT), yang membangun model klasifikasi berdasarkan atribut-atribut data secara bertahap dari akar hingga daun (Quinlan, 1986) . Pengembangan dari pendekatan tersebut menghasilkan algoritma C4.5, yang mampu menangani atribut numerik maupun kategorikal serta menghasilkan aturan keputusan (*rules*) yang mudah dipahami oleh pengguna.

Metode C4.5 bekerja melalui proses menentukan besaran entropi serta perolehan informasi guna menentukan variabel yang paling tepat diterapkan sebagai klasifikasi. Proses ini dilakukan secara berulang hingga seluruh data terklasifikasi ke dalam kelas tertentu. Penelitian sebelumnya menunjukkan bahwa algoritma C4.5 mampu menghasilkan model klasifikasi status gizi balita dengan tingkat akurasi yang baik, salah satunya mencapai akurasi 80% berdasarkan pengujian *confusion matrix* (Matondang et al., 2024). Hasil tersebut menunjukkan bahwa C4.5 layak digunakan sebagai metode dalam membantu proses identifikasi status gizi secara objektif.

Meskipun penelitian terkait klasifikasi status gizi menggunakan algoritma C4.5 telah dilakukan di beberapa wilayah, penerapan khusus untuk mengklasifikasikan risiko gizi buruk pada balita dengan studi kasus di UPTD Puskesmas Gunung Meriah belum dilakukan. Setiap wilayah memiliki karakteristik data dan kondisi masyarakat yang berbeda, sehingga diperlukan model klasifikasi yang dibangun berdasarkan data lokal agar hasilnya lebih relevan dan aplikatif.

Berdasarkan permasalahan tersebut, diperlukan suatu sistem klasifikasi yang mampu mengolah data balita secara terstruktur dan objektif untuk menentukan risiko gizi buruk menggunakan metode C4.5. Dengan adanya sistem ini, diharapkan tenaga kesehatan dapat lebih cepat dalam melakukan identifikasi dini serta menentukan prioritas penanganan balita yang berisiko.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana memanfaatkan data antropometri balita yang diperoleh dari UPTD Puskesmas Gunung meriah untuk mengidentifikasi risiko gizi buruk?
2. Bagaimana menerapkan metode C4.5 dalam mengidentifikasi risiko gizi buruk pada balita berdasarkan atribut seperti usia, jenis kelamin, berat badan, tinggi badan, dan indikator status gizi?
3. Bagaimana tingkat akurasi model klasifikasi yang dihasilkan oleh algoritma C4.5 dalam memprediksi risiko gizi buruk pada balita?
4. Atribut apa yang paling berpengaruh dalam menentukan klasifikasi risiko gizi buruk berdasarkan hasil pembentukan pohon keputusan?

1.3. Batasan Masalah

Agar penelitian lebih terarah dan tidak menyimpang dari tujuan, maka batasan masalah dalam penelitian ini adalah:

1. Penelitian ini menggunakan data balita yang diperoleh dari UPTD Puskesmas Gunung Meriah periode Oktober 2025 - Januari 2026.
2. Data yang digunakan meliputi atribut antropometri dan informasi dasar balita, seperti usia, jenis kelamin, berat badan, tinggi badan, serta indikator status gizi yang tersedia pada data posyandu.
3. Klasifikasi yang dilakukan berfokus pada penentuan risiko gizi buruk pada balita berdasarkan data yang tersedia.
4. Penelitian ini tidak membahas faktor eksternal seperti kondisi sosial ekonomi, pola makan, penyakit penyerta, maupun faktor lingkungan yang tidak tercantum dalam *dataset*.

1.4. Tujuan Penelitian

Tujuan dari penelitian adalah:

1. Memanfaatkan data antropometri balita dari UPTD Puskesmas Gunung meriah untuk mengidentifikasi pola risiko gizi buruk secara sistematis
2. Menerapkan metode C4.5 dalam membangun model klasifikasi risiko gizi buruk pada balita berdasarkan atribut usia, jenis kelamin, berat badan, dan tinggi badan
3. Mengukur tingkat akurasi model klasifikasi C4.5 dalam memprediksi risiko gizi buruk menggunakan pengujian *confusion matrix*
4. Mengidentifikasi atribut yang paling berpengaruh dalam menentukan klasifikasi risiko gizi buruk berdasarkan hasil pembentukan pohon keputusan

1.5. Manfaat Penelitian

1. Bagi Penulis:
 - a. Menambah wawasan dan pengalaman dalam penerapan metode data Mining
 - b. Mengembangkan kemampuan analisis data kesehatan
 - c. Menjadi sarana penerapan ilmu yang diperoleh selama perkuliahan
2. Bagi Instansi (UPTD Puskesmas Gunung Meriah)
 - a. Mengidentifikasi balita yang berisiko mengalami gizi buruk secara lebih cepat
 - b. Mendukung pengambilan keputusan dalam penanganan masalah gizi
 - c. Memanfaatkan data posyandu lebih optimal
 - d. Menentukan prioritas intervensi terhadap balita yang membutuhkan perhatian khusus

3. Bagi Keilmuan

- a. Memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya pada bidang *data mining* dan sistem informasi kesehatan
- b. Menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan penerapan algoritma klasifikasi, khususnya metode C4.5, dalam analisis data kesehatan balita

BAB II

LANDASAN TEORI

2.1. Gizi Bayi dan Risiko Gizi Buruk

Gizi buruk merupakan masalah yang perlu penanganan serius. Berbagai upaya telah dilakukan pemerintah antara lain melalui revitalisasi posyandu dalam meningkatkan cakupan penimbangan balita, penyuluhan dan pendampingan, pemberian Makanan Pendamping ASI (MP-ASI) atau Pemberian Makanan Tambahan (PMT), peningkatan akses dan mutu pelayanan gizi melalui tata laksana gizi buruk di Puskesmas Perawatan dan Rumah Sakit, penanggulangan penyakit menular dan pemberdayaan masyarakat melalui Keluarga Sadar Gizi (Ariana, 2010)

Gizi buruk atau malnutrisi merupakan masalah kesehatan yang sering dialami balita dan dapat disebabkan oleh faktor internal maupun eksternal. Jika tidak dicegah dan ditangani sejak dini, kondisi ini dapat mengganggu kesehatan, pertumbuhan, dan perkembangan anak serta menurunkan produktivitasnya saat dewasa (Lestari, 2022) . Menurut *World Health Organization*, gizi buruk pada anak merupakan salah satu penyebab utama meningkatnya risiko kematian, gangguan perkembangan kognitif, serta berbagai penyakit kronis di masa depan (World Health Organization, 2021).

Dalam pelayanan kesehatan masyarakat, pemantauan status gizi bayi dilakukan melalui pengukuran antropometri seperti berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), dan berat badan menurut tinggi badan (BB/TB). Data tersebut digunakan oleh tenaga kesehatan untuk menilai

kondisi gizi bayi serta menentukan intervensi yang tepat guna mencegah terjadinya gizi buruk. Oleh karena itu, pemahaman mengenai kondisi bayi menjadi sangat penting sebagai dasar dalam penentuan kelayakan pemberian makanan pendamping asi (MP-ASI) bagi bayi yang berisiko mengalami masalah gizi.

2.2. Konsep Dasar Decision Tree

Decision Tree merupakan metode yang menggambarkan alur pengambilan keputusan dalam bentuk pohon. Struktur ini terdiri dari node yang merepresentasikan suatu keputusan dan cabang yang menunjukkan kemungkinan hasil dari keputusan tersebut (Feby, n.d.). Teknik tersebut beroperasi melalui proses memisahkan data dengan berangsur-angsur menjadi golongan berukuran kecil sampai didapatkan keluaran pengelompokan yang seragam. Struktur *Decision Tree* terdiri dari *root node* sebagai atribut utama, *internal node* sebagai pengujian atribut, cabang sebagai hasil pengujian, dan *leaf node* sebagai keputusan akhir atau kelas. Metode ini merupakan bagian dari supervised learning yang memanfaatkan data latih (*training dataset*) sebagai acuan untuk membangun model pengambilan keputusan (Kurniawan, 2020).

Pendekatan pembelajaran *Decision Tree* bersifat *top-down* dan menggunakan strategi *greedy*, yaitu memilih atribut terbaik pada setiap langkah untuk memisahkan data. Pemilihan atribut biasanya didasarkan pada ukuran tertentu seperti informasi *gain*, *gain ratio*, atau *gain index*, tergantung pada algoritma yang digunakan. Tujuan dari pemilihan atribut ini adalah untuk menghasilkan pembagian data yang paling informatif sehingga proses klasifikasi menjadi lebih akurat.

Pembentukan model Decision Tree dapat dilakukan dengan menggunakan beberapa algoritma pembelajaran, salah satunya adalah CART (*Classification and Regression Trees*) yang diterapkan dalam Scikit-learn. Algoritma ini memiliki hubungan dengan pengembangan metode C4.5 dan ID3. Sementara itu, ID3 (*Iterative Dichotomiser*) merupakan algoritma yang diperkenalkan oleh Ross Quinlan, seorang peneliti di bidang *Machine Learning* (Puspitorini & Sintawati, 2021).

Pengembangan dari ID3 adalah C4.5 yang mampu menangani atribut kontinu, *missing value*, serta menggunakan *gain ratio* untuk mengurangi bias terhadap atribut dengan banyak nilai. Selanjutnya, C5.0 merupakan versi yang lebih efisien dari C4.5 dengan kecepatan pemrosesan yang lebih tinggi dan penggunaan memori yang lebih rendah, sehingga cocok untuk *dataset* berukuran besar.

Selain itu, terdapat algoritma lain seperti CART yang membangun pohon biner menggunakan *gain index* sebagai ukuran *impurity*, serta CHAID yang menggunakan uji *Chi-Square* untuk menentukan pembagian data. Algoritma-algoritma tersebut memiliki kelebihan dan kekurangan masing-masing tergantung pada karakteristik data yang digunakan.

Decision Tree banyak digunakan dalam berbagai bidang karena kemampuannya yang mudah dipahami dan diinterpretasikan. Metode ini diterapkan dalam bisnis, kesehatan, deteksi penipuan, *e-commerce*, serta berbagai sistem pendukung keputusan lainnya. Namun demikian, *Decision Tree* juga memiliki beberapa kelemahan, seperti kemungkinan terjadinya *overfitting*, fragmentasi data, serta kesulitan dalam menangani hubungan data yang kompleks (Patel & Rana, 2014).

2.3. Algoritma C4.5

Salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan (*decision tree*) adalah algoritma C4.5. Algoritma C4.5 merupakan algoritma yang sangat populer yang digunakan oleh banyak peneliti di dunia, hal ini dijelaskan oleh Xindong Wu dan Vipin Kumar dalam bukunya yang berjudul *The Top Ten Algorithms in Data Mining*. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang diciptakan oleh J. Rose Quinlan (Mardi, 2017).

Algoritma C4.5 bekerja dengan membangun pohon keputusan berdasarkan atribut yang memiliki nilai informasi tertinggi terhadap kelas target. Proses pemilihan atribut dilakukan menggunakan konsep *entropy* dan *information gain ratio*. *Entropy* digunakan untuk mengukur tingkat ketidakpastian suatu data, sedangkan *gain ratio* digunakan untuk menentukan atribut yang akan dijadikan *node* pada pohon keputusan

Secara umum, C4.5 memiliki beberapa keunggulan dibandingkan metode klasifikasi lainnya, antara lain mampu menangani data kategorikal maupun numerik, dapat mengatasi *missing value*, serta menghasilkan model yang mudah dipahami oleh pengguna. Karena karakteristik tersebut, C4.5 banyak digunakan dalam berbagai bidang seperti kesehatan, pendidikan, keuangan, dan sistem pendukung keputusan.

Dalam bidang kesehatan, algoritma C4.5 sering digunakan untuk membantu proses pengambilan keputusan berbasis data. Penelitian oleh sari dkk. (2021) menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan data kesehatan secara efektif berdasarkan sejumlah parameter medis, sehingga dapat membantu tenaga kesehatan dalam menentukan tindakan yang tepat secara lebih cepat dan objektif.

Selain itu, penelitian lain yang dilakukan oleh Prasetyo (2012) menyatakan bahwa algoritma C4.5 memiliki tingkat akurasi yang tinggi dalam proses klasifikasi karena menggunakan konsep *gain ratio* yang mampu mengurangi bias pemilihan atribut dengan banyak nilai. Hal ini menjadikan C4.5 sebagai salah satu algoritma *decision tree* yang paling populer dalam *data mining*.

Dengan kemampuan tersebut, algoritma C4.5 sangat sesuai digunakan untuk mengklasifikasikan kelayakan makanan pada bayi berisiko gizi buruk, karena dapat mengolah berbagai atribut seperti usia bayi, berat badan bayi, tinggi badan bayi, serta kondisi kesehatan untuk menghasilkan keputusan yang akurat dan mudah dipahami.

2.3.1. Entropy

Entropy merupakan ukuran ketidakpastian atau tingkat *impurity* suatu kumpulan data. Semakin tinggi nilai *entropy*, semakin heterogen data tersebut. Entropi dimanfaatkan guna mengetahui sejauh mana efektivitas sebuah variabel untuk membagi kumpulan data menjadi kelompok berlainan. Menurut quinlan (1993), *entropy* bernilai nol apabila semua data berada dalam satu kelas yang sama (homogen), dan akan meningkat seiring dengan meningkatnya keragaman kelas dalam *dataset*. Rumus *entropy* adalah:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan :

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi dari S_i terhadap S

2.3.2. Mencari Nilai Gain

Information Gain digunakan untuk menentukan atribut mana yang paling baik digunakan sebagai pemisah data pada setiap *node*. *Information Gain* mengukur pengurangan *entropy* setelah *dataset* dibagi berdasarkan suatu atribut. Rumus *Information Gain* adalah:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S = himpunan kasus

A = atribut

N = jumlah partisi atribut A

[S_i] = jumlah kasus pada partisi ke-i

[S] = jumlah kasus pada S

2.4. Perangkat Lunak (*Software*)

Merupakan sekumpulan program, data, dan instruksi yang memberi tahu komputer atau perangkat elektronik apa yang harus dilakukan.

2.4.1. Python

Diperkenalkan pada Desember 1989 oleh Guido Van, Python merupakan satu di antara bahasa komputer general level tinggi yang paling dikenal. Python merupakan bahasa komputer yang menggunakan sistem pengetikan dinamis, kerap dikenal sebagai duck typing dalam pengembangan perangkat lunak (Grinberg, 2018). Bahasa ini merupakan bahasa interpretatif. Sebab, bahasa tersebut tidak membutuhkan proses kompilasi tersendiri, tetapi diterjemahkan menjadi kode byte serta dijalankan secara langsung. Salah satu faktor kepopuleran Python disebabkan oleh kemudahan untuk dibaca. Sebab, instruksi program yang

mudah dipahami merupakan harapan bagi seorang insinyur perangkat lunak. Hal tersebut penting karena pengembang perlu mampu mengerti makna kode tanpa bergantung pada penulisnya. Kemudahan membaca kode sebenarnya bergantung pada pembuatnya, tetapi Python lebih unggul dalam aspek tersebut melalui pendekatan yang menyerupai bahasa manusia. Bahasa ini juga dikenal sebagai metode sederhana untuk mempelajari bahasa komputer. Hal ini sangat memudahkan pemrogram dalam memangkas durasi yang digunakan untuk memahami bahasa komputer lainnya (Chandra, 2023).

2.4.2. Python Flask

Python Flask merupakan kerangka kerja yang tergolong ringan berdasarkan mayoritas tolok ukur kerangka kerja. Ukurannya tergolong cukup untuk dikategorikan sebagai kerangka kerja mikro dan relatif ringkas sehingga pengguna yang telah mengenal Python Flask dapat membaca serta mengerti seluruh kode sumbernya. Walaupun tampak ringkas, hal tersebut tidak berarti Python Flask dapat dinilai lebih rendah daripada kerangka kerja lainnya. Python Flask dirancang sebagai kerangka kerja mikro yang bisa dikembangkan secara menyeluruh. Hal tersebut menunjukkan bahwa Flask mempunyai fondasi kuat pada fungsi-fungsi dasarnya, sedangkan ekstensi yang tersedia melengkapi bagian lainnya. Pengguna juga dapat menentukan kumpulan ekstensi sesuai kebutuhan. Dengan demikian, pengguna tidak mengalami kelebihan ukuran berkas yang dapat memperbesar ukuran berkas yang dimiliki.

Python Flask memiliki sejumlah ketergantungan inti sebagai pendukungnya, yakni pengaturan rute, penelusuran kesalahan, dan Web Server Gateway Interface (WSGI). Bagian sistem Python Flask bersumber dari Werkzeug. Dukungan

template yang tersedia ditunjang oleh Jinja2, sedangkan integrasi baris perintah bersumber dari Click (Grinberg, 2018). Seluruh ketergantungan tersebut merupakan komponen inti Python Flask. Armin Ronacher selaku pembuat Python Flask tidak menyertakan ketergantungan bawaan untuk sistem basis data, validasi formulir web, autentikasi pengguna, serta tugas tingkat tinggi lainnya. Fungsi-fungsi tersebut umumnya dipadukan sebagai ketergantungan inti pada kerangka kerja lain. Sebagai pengembang, kita tentu mempunyai keraguan ketika menentukan ekstensi yang sesuai atau ideal untuk proyek yang dikembangkan. Bahkan, kita dapat mengembangkan ekstensi sendiri jika dibutuhkan. Hal tersebut tentunya bertolak belakang dengan kerangka kerja lainnya.

2.4.3. Scikit-Learn

Penggunaan *Scikit-learn* dalam penelitian memberikan kemudahan dalam mengimplementasikan berbagai algoritma *machine learning* dengan struktur kode yang sistematis dan terstandarisasi. Melalui penerapan langkah-langkah pengenalan persoalan, penentuan metode, penyiapan data, serta penerapan model, rangkaian kegiatan pengembangan sistem pembelajaran mesin bisa dilaksanakan dengan terstruktur dan selaras terhadap sasaran riset

Implementasi ini menunjukkan bahwa *scikit-learn* tidak hanya berfungsi sebagai alat pemrograman, tetapi juga sebagai *framework* konseptual dalam membangun model pembelajaran mesin berbasis python (Fahmi, 2023)

2.4.4. Bootstrap

Bootstrap adalah *framework front-end* yang digunakan untuk mempermudah proses perancangan dan pengembangan antarmuka (*interface*) *website*. *Framework* ini menyediakan kumpulan komponen siap pakai seperti *grid system*,

tombol, formulir, navigasi, hingga elemen interaktif, sehingga pengembang tidak perlu membuat desain dari nol.

Salah satu keunggulan utama *bootstrap* adalah kemampuannya dalam membangun tampilan *website* yang responsif. Dengan sistem *grid* yang fleksibel, halaman *web* dapat menyesuaikan tata letak secara otomatis pada berbagai ukuran layar, baik desktop, tablet, maupun perangkat *mobile*. Hal ini sangat penting dalam pengembangan web modern yang menuntut akses lintas perangkat.

Bootstrap juga dirancang agar mudah digunakan, bahkan oleh pemula. Struktur kelas (*class*) yang konsisten memungkinkan pengembang untuk menerapkan gaya visual hanya dengan menambahkan atribut tertentu pada elemen HTML. Selain itu, *Bootstrap* juga menyediakan dokumentasi lengkap serta standar desain yang rapi dan modern.

Dengan berbagai kemudahan tersebut, *Bootstrap* banyak digunakan untuk mempercepat proses pengembangan *web*, meningkatkan konsisten tampilan, serta memastikan kompatibilitas antar perangkat dan browser (Mardi Yudhi Putra et al., 2021).

2.5. Unified Modeling Language (UML)

Unified Modeling Language (UML) merupakan bahasa pemodelan standar yang digunakan dalam rekayasa perangkat lunak untuk memvisualisasikan, merancang, dan mendokumentasikan sistem. UML membantu pengembangan dalam menggambarkan struktur dan perilaku sistem secara jelas sehingga memudahkan proses analisis dan pengembangan sistem yang akan dibangun.

UML terdiri dari berbagai jenis diagram yang memiliki fungsi berbeda, seperti *Use Case Diagram*, *Activity Diagram*, *Class Diagram*, dan *Sequence Diagram*.

Setiap diagram digunakan untuk merepresentasikan aspek tertentu dari sistem, baik dari sisi struktur maupun proses.

Dalam penelitian ini, UML bertujuan untuk memodelkan sistem klasifikasi risiko gizi buruk pada balita berbasis web. Diagram yang digunakan meliputi *Use Case Diagram*, *Sequence Diagram*, *Class Diagram* dan *Activity Diagram*, karena keempat diagram tersebut mampu menggambarkan interaksi pengguna dengan sistem serta alur proses yang terjadi secara sistematis.

2.5.1. Use Case Diagram

Use Case Diagram merupakan diagram yang digunakan untuk menggambarkan hubungan antara aktor dengan sistem. Diagram ini menunjukkan fungsi-fungsi utama yang dapat dilakukan oleh pengguna terhadap sistem yang dibangun.

Aktor adalah entitas yang berinteraksi dengan sistem, yang dapat berupa manusia atau sistem lain. Sedangkan *Use Case* merupakan layanan atau proses yang disediakan oleh sistem untuk aktor. Dengan adanya *Use Case Diagram*, pengembang dapat memahami kebutuhan sistem dari sudut pandang pengguna serta mengetahui batasan sistem secara keseluruhan.

Dalam konteks penelitian ini, *Use Case Diagram* digunakan untuk menggambarkan interaksi antara pengguna (petugas kesehatan) dengan sistem klasifikasi risiko gizi buruk. Beberapa aktivitas yang dapat dilakukan antara lain mengelola data balita, melakukan proses klasifikasi menggunakan metode C4.5, serta melihat hasil klasifikasi.

2.5.2. Activity Diagram

Activity Diagram merupakan diagram yang digunakan untuk menggambarkan alur aktivitas atau proses kerja dalam suatu sistem. Diagram ini menunjukkan urutan kegiatan dari awal hingga akhir, termasuk percabangan keputusan (*decision*), proses paralel, serta aliran kontrol dalam sistem.

Activity Diagram membantu dalam memahami logika proses bisnis atau sistem secara lebih rinci, sehingga memudahkan dalam perancangan dan implementasi sistem. Dengan menggunakan diagram ini, pengembang dapat melihat bagaimana suatu proses berjalan dan bagaimana interaksi antar komponen dalam sistem terjadi.

Dalam penelitian ini, *Activiry* Diagram digunakan untuk menggambarkan alur proses sistem klasifikasi risiko gizi buruk, mulai dari *input* data balita, proses pengolahan data menggunakan algoritma C4.5, hingga menghasilkan *output* berupa klasifikasi risiko gizi buruk.

Dengan adanya pemodelan menggunakan UML, diharapkan proses perancangan sistem dapat dilakukan secara lebih terstruktur dan mudah dipahami, sehingga sistem yang dibangun sesuai dengan kebutuhan pengguna.

2.6. Literatur Review

Berikut Adalah tinjauan pustaka yang relevan pada penelitian seperti pada tabel di bawah ini:

Tabel 2.1. Literatur *Review*

No	Peneliti Dan Tahun	Judul Penelitian	Kelebihan	Kekurangan
1.	Firda Maisa Hana (2020)	Klasifikasi Penderita Penyakit Diabetes Menggunakan algoritma	Akurasi sanat tinggi, menggunakan <i>confusion matrix</i> lengkap, model mudah dipahami	Tidak berbasis sistem, studi hanya pada bidang kesehatan

		<i>decision Tree</i> C4.5		
2.	Yuli Mardi (2017)	Data Mining: klasifikasi Menggunakan Algoritma c4.5	Landasan teori kuat, tahapan penelitian sistematis	Tidak berbasis studi kasus spesifik, tidak menampilkan hasil yang akurat
3.	Yusuf Ramadhan Nasution (2024)	Penerapan Algoritma c4.5 Pada Klasifikasi Status Gizi Balita	<i>Dataset</i> cukup besar, pembagian data jelas, analisis atribut berpengaruh	Akurasi masih sedang, belum berbasis web
4.	Riska Rofiani dkk (2024)	Prediksi Kanker Paru-Paru menggunakan Algoritma c4.5	Mengidentifikasi atribut dominan, <i>dataset</i> nyata (kaggle)	<i>Dataset</i> relatif kecil (150 data), tidak implementasi sistem
5.	Sutan faisal (2019)	Klasifikasi Kepuasan Pelanggan Sewa Kamera Menggunakan AlgoritmaC4.5	Evaluasi model lengkap, akurasi tinggi, studi kasus bisnis	Tidak fokus pada bidang kesehatan, tidak berbasis sistem web
6	(Saleh, 2019)	Klasifikasi Status Gizi Balita Menggunakan Metode K- <i>Nearest Neighbor</i>	Menjelaskan 4 kategori status gizi, menggunakan data antropometri (U, BB, TB, LK), perhitungan Euclidean Distance jelas,	<i>dataset</i> kecil tidak menggunakan LILA, tidak ada akurasi
7	(Hafizan & Putri, 2020)	Penerapan Metode Klasifikasi <i>Decision Tree</i> Pada Status Gizi Balita Di Kabupaten Simalungun	Membahas konsep <i>Decision Tree</i> dan C4.5 secara jelas,	hanya 3 atribut, tidak implementasi sistem
8	(Yoshe & Hadikurniawati, 2021)	Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada	menggunakan <i>confusion matrix</i> , berbasis web dengan fitur <i>import</i> data dan proses klasifikasi	fokus pada stunting, pembagian data latih dan uji tidak seimbang

		Balita		
9	(Islam, 2022)	Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita	klasifikasi 6 kategori gizi, berbasis web	hanya menggunakan indeks BB/TB tanpa mempertimbangkan BB/U dan TB/U
10	(Nasution, 2025)	Faktor Risiko Terjadinya Stunting pada Balita	Membahas faktor risiko stunting dari 3 aspek, menjelaskan dampak jangka pendek dan panjang, rekomendasi pencegahan jelas	tidak ada metode data <i>mining</i> /C4.5, tidak berbasis web

BAB III

ANALISA DAN PERANCANGAN SISTEM

3.1. Analisis Permasalahan

Permasalahan gizi buruk pada balita masih menjadi salah satu isu peting dalam bidang kesehatan masyarakat. Balita yang mengalami kekurangan gizi berisiko mengalami gangguan pertumbuhan fisik, perkembangan kognitif, serta meningkatkan kemungkinan terkena berbagai penyakit pada masa mendatang. Oleh karena itu, pemantauan status gizi balita secara rutin menjadi langkah penting dalam upaya pencegahan dan penanganan masalah gizi di masyarakat.

UPTD Puskesmas Gunung Meriah sebagai salah satu fasilitas pelayanan kesehatan tingkat pertama memiliki tanggung jawab dalam melakukan pemantauan kondisi kesehatan balita di wilayah kerjanya. Setiap periode pelayanan, puskesmas mengumpulkan berbagai data balita yang meliputi umur, berat badan, tinggi badan, serta hasil penilaian status gizi yang diperoleh dari kegiatan posyandu maupun pemeriksaan kesehatan rutin. Data tersebut terus bertambah seiring dengan meningkatnya jumlah balita yang melakukan pemeriksaan.

Namun, dalam praktiknya, proses identifikasi balita yang berisiko mengalami gizi buruk masih dilakukan secara manual dengan cara mengevaluasi data satu per satu berdasarkan parameter tertentu. Pendekatan ini memiliki beberapa keterbatasan, terutama ketika jumlah data yang dikelola semakin besar. Proses analisis menjadi lebih lama dan berpotensi menimbulkan inkonsistensi dalam penilaian karena sangat bergantung pada interpretasi tenaga kesehatan.

Selain itu, data yang telah dikumpulkan belum dimanfaatkan secara optimal untuk menemukan pola-pola penting yang dapat membantu dalam proses pengambilan keputusan.

Di sisi lain, perkembangan teknologi informasi telah menghadirkan berbagai metode analisis data yang mampu membantu mengolah data dalam jumlah besar secara lebih sistematis. Salah satu pendekatan yang dapat digunakan adalah teknik *data Mining*, yaitu proses pengolahan data untuk menemukan pola atau informasi yang bermanfaat dari kumpulan data yang besar. Dalam konteks penelitian ini, teknik klasifikasi dalam *data Mining* dapat digunakan untuk mengelompokkan data balita berdasarkan tingkat risiko gizi buruk.

Metode yang digunakan dalam penelitian ini adalah algoritma C4.5 yang merupakan salah satu algoritma *decision tree* yang populer dalam proses klasifikasi data. Algoritma ini bekerja dengan membangun pohon keputusan berdasarkan atribut yang paling berpengaruh terhadap kelas target melalui perhitungan nilai *entropy* dan *information gain*. Dengan menggunakan metode ini, sistem dapat menghasilkan model klasifikasi yang mampu mengidentifikasi risiko gizi buruk pada balita secara lebih objektif dan sistematis.

Berdasarkan permasalahan tersebut, diperlukan suatu sistem yang mampu mengolah data balita secara terstruktur dan memanfaatkan teknik klasifikasi untuk membantu tenaga kesehatan dalam mengidentifikasi risiko gizi buruk. Dengan adanya sistem klasifikasi berbasis metode C4.5, diharapkan proses analisis data balita dapat dilakukan secara lebih cepat, akurat, dan konsisten sehingga dapat membantu pihak puskesmas dalam menentukan langkah penanganan yang tepat terhadap balita yang berisiko mengalami gizi buruk.

3.2. Algoritma Sistem

Algoritma sistem menjelaskan urutan proses yang digunakan untuk melakukan klasifikasi risiko gizi buruk pada balita menggunakan algoritma C4.5. Data yang digunakan terdiri dari atribut umur, jenis kelamin, berat badan, tinggi badan, dan LILA, sedangkan status gizi digunakan sebagai target klasifikasi.

Tahapan algoritma C4.5 dalam penelitian ini meliputi :

1. Menyiapkan *dataset* balita sebagai data *training*.
2. Menentukan atribut yang digunakan dalam proses klasifikasi, yaitu umur, jenis kelamin, berat badan, tinggi badan, dan LILA.
3. Menentukan kelas target berupa gizi baik, berisiko gizi buruk, dan gizi buruk.
4. Menghitung *entropy* dataset.
5. Menghitung *information gain* dari setiap atribut.
6. Menghitung *split information* pada setiap atribut.
7. Menghitung *gain ratio* untuk menentukan atribut terbaik.
8. Memilih atribut dengan nilai *gain ratio* tertinggi sebagai *root node*.
9. Membentuk cabang dan menghitung kembali nilai *entropy*, *gain*, *split information*, dan *gain ratio* pada setiap cabang yang masih memiliki lebih dari satu kelas.
10. Mengulangi proses tersebut sampai terbentuk pohon keputusan.
11. Menghasilkan aturan keputusan yang digunakan untuk menentukan klasifikasi risiko gizi buruk pada balita.

3.2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data yang diperoleh dari UPTD Puskesmas Gunung Meriah, dan merupakan data status gizi balita periode

2025-2026. *Dataset* yang diperoleh mencakup beberapa atribut utama, seperti umur, berat badan, tinggi badan dan jenis kelamin sebagai variable target. Data tersebut digunakan sebagai data latih (*training data*) dalam proses klasifikasi menggunakan algoritma C4.5.

Tabel 3.1. Sumber Data

No	Atribut	Tipe Data	Peran
1	Umur	Numerik	Atribut
2	Jenis Kelamin	Kategorikal	Atribut
3	Berat Badan	Numerik	Atribut
4	Tinggi Badan	Numerik	Atribut
5	LILA	Numerik	Atribut
6	Status Gizi	Kategorikal	Target

Keterangan:

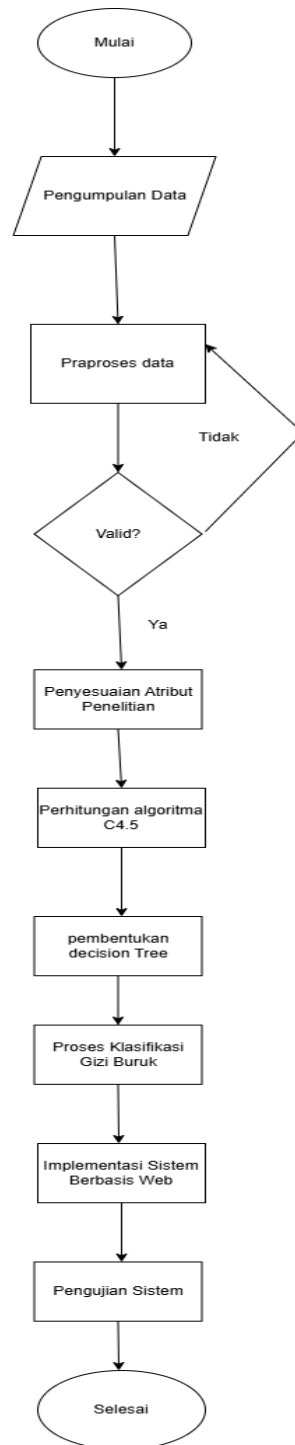
1. Umur, yaitu usia balita yang digunakan sebagai atribut dalam proses klasifikasi
2. Jenis Kelamin, menggambarkan gender anak usia bawah lima tahun yang mencakup pria serta wanita.
3. Berat Badan, merupakan bobot tubuh anak usia bawah lima tahun dengan ukuran kilogram (kg).
4. Tinggi Badan, merupakan tinggi tubuh anak usia bawah lima tahun dengan ukuran sentimeter (cm).
5. LILA, adalah pengukuran keliling bagian atas lengan yang dimanfaatkan selaku satu di antara variabel pengukuran dimensi tubuh pada tahapan pengelompokan kondisi gizi anak usia bawah lima tahun.

6. Status Gizi, merupakan kelas target yang digunakan dalam proses klasifikasi, seperti gizi baik, gizi kurang, atau gizi buruk.

3.3. Tahapan Penelitian

Tahapan penelitian merupakan rangkaian langkah sistematis yang dilakukan untuk mencapai tujuan penelitian yang telah ditetapkan. Dalam penelitian ini, tahapan dilakukan secara terstruktur mulai dari identifikasi masalah hingga penarikan Kesimpulan. Setiap tahapan dirancang agar proses klasifikasi Risiko gizi buruk pada balita menggunakan metode C4.5 dapat berjalan secara optimal dan menghasilkan sistem yang akurat serta dapat diimplementasikan dengan baik.

Alur tahapan penelitian ini ditujukan pada gambar *flowchart* tahapan penelitian. *Flowchart* tersebut menggambarkan proses penelitian yang dimulai dari pengumpulan data hingga implementasi sistem berbasis web.



Gambar 3.1. *Flowchart Tahapan Penelitian*

3.3.1. Pengumpulan Data

Tahapan pengumpulan data dilakukan untuk memperoleh data yang relevan dan dibutuhkan dalam penelitian. Data yang digunakan dalam penelitian ini diperoleh dari UPTD Puskesmas Gunung Meriah. Data tersebut berupa data balita yang meliputi beberapa atribut, yaitu umur, jenis kelamin, berat badan, tinggi badan, serta status gizi.

Proses pengumpulan data dilakukan dengan cara dokumentasi terhadap data yang telah tersedia. Data yang diperoleh kemudian dikumpulkan dan disusun dalam bentuk *dataset* yang akan digunakan dalam proses pengolahan selanjutnya. Kualitas data yang baik sangat berpengaruh terhadap hasil klasifikasi yang akan dihasilkan.

3.3.2. Praproses Data

Data yang telah dikumpulkan tidak selalu dalam kondisi siap digunakan, sehingga diperlukan tahap praproses data. Tahap ini bertujuan untuk meningkatkan kualitas data agar lebih bersih, konsisten, dan sesuai untuk digunakan dalam proses klasifikasi

Adapun proses praproses yang dilakukan meliputi:

1. *Data Cleaning*, yaitu proses pembersihan data dari nilai tidak valid, duplikasi data, atau kesalahan *input*.
2. Penanganan *Missing Value*, yaitu mengatasi data yang kosong atau tidak lengkap agar tidak mengganggu proses analisis.
3. Transformasi Data, yaitu mengubah data ke dalam format yang sesuai jika diperlukan, misalnya normalisasi atau pengelompokan data tertentu.

Tahapan ini sangat penting karena kualitas data akan sangat mempengaruhi performa algoritma C4.5 dalam menghasilkan model klasifikasi.

3.3.3. Penentuan Atribut dan Target

Pada tahap ini, dilakukan penentuan variabel yang digunakan dalam proses klasifikasi. Variabel dibagi menjadi dua jenis, yaitu atribut (variabel *input*) dan target (variabel *output*).

Atribut yang digunakan dalam penelitian ini meliputi umur, jenis kelamin, berat badan, dan tinggi badan. Sedangkan variabel target adalah status gizi balita yang akan diklasifikasikan ke dalam beberapa kategori, seperti gizi baik, gizi kurang, dan gizi buruk.

Penentuan atribut dan target ini bertujuan untuk memastikan bahwa proses klasifikasi berjalan sesuai dengan tujuan penelitian, yaitu memprediksi status gizi berdasarkan data yang tersedia.

3.3.4. Penerapan Algoritma C4.5

Pada tahap ini dilakukan proses utama penelitian, yaitu penerapan algoritma C4.5 untuk membangun model klasifikasi. Algoritma C4.5 merupakan salah satu metode dalam *data mining* yang digunakan untuk membentuk pohon keputusan (*decision tree*) berdasarkan data yang tersedia.

Proses dalam algoritma C4.5 meliputi beberapa langkah, yaitu:

1. Menghitung nilai *entropy* untuk setiap atribut.
2. Menghitung nilai *information gain*.
3. Menentukan atribut terbaik sebagai akar (*root*) pohon keputusan.
4. Membagi data berdasarkan atribut terpilih.
5. Mengulangi proses hingga terbentuk pohon keputusan yang optimal.

Hasil pada tahap ini berupa model *decision tree* yang dapat digunakan untuk menghasilkan aturan-aturan klasifikasi (*rules*) dalam menentukan status gizi balita.

3.3.5. Implementasi Sistem

Setelah model klasifikasi diperoleh, tahap selanjutnya adalah implementasi ke dalam sistem berbasis web. Sistem ini dirancang untuk mempermudah pengguna, seperti tenaga kesehatan, dalam melakukan proses klasifikasi data balita secara cepat dan mudah.

Pada sistem ini, pengguna dapat memasukkan data balita, kemudian sistem akan memproses data tersebut menggunakan model C4.5 dan menghasilkan *output* berupa klasifikasi status gizi. Implementasi sistem ini diharapkan dapat meningkatkan efisiensi dan akurasi dalam pengambilan keputusan.

3.3.6. Pengujian Sistem

Tahap pengujian dilakukan untuk mengetahui kinerja sistem yang telah dibangun. Pengujian bertujuan untuk memastikan bahwa sistem dapat berjalan dengan baik dan menghasilkan *output* yang sesuai dengan harapan.

Pengujian dapat dilakukan dengan menggunakan data uji yang telah disiapkan sebelumnya. Hasil klasifikasi dari sistem kemudian dibandingkan dengan data sebenarnya untuk mengetahui tingkat akurasi. Selain itu, pengujian juga dilakukan untuk memastikan bahwa semua fitur dalam sistem berjalan dengan baik.

3.3.7. Evaluasi dan Analisis Hasil

Pada tahap ini dilakukan evaluasi terhadap hasil yang diperoleh dari pengujian sistem. Evaluasi bertujuan untuk mengetahui tingkat keberhasilan model klasifikasi yang telah dibangun.

Analisis dilakukan dengan melihat tingkat akurasi, kelebihan, serta keterbatasan dari sistem. Hasil evaluasi kemudian digunakan sebagai dasar untuk

menentukan apakah sistem sudah layak digunakan atau masih perlu dilakukan perbaikan.

3.3.8. Penarikan Kesimpulan

Tahap terakhir dalam penelitian adalah penarikan kesimpulan. Kesimpulan diperoleh berdasarkan seluruh proses yang telah dilakukan, mulai dari pengolahan data hingga evaluasi hasil.

Pada tahap ini juga dapat diberikan saran untuk pengembangan penelitian selanjutnya, sehingga penelitian ini menjadi referensi untuk penelitian yang lebih baik di masa yang akan datang.

3.4. Perhitungan Manual Algoritma C4.5

Pada subbab ini dilakukan perhitungan manual algoritma *Decision Tree* C4.5 menggunakan *dataset* balita yang digunakan dalam penelitian. Kalkulasi dilaksanakan guna mendapatkan struktur keputusan yang bisa dimanfaatkan untuk mengelompokkan kondisi gizi anak usia bawah lima tahun menjadi tiga kelompok, yakni kondisi gizi baik, berpotensi mengalami gizi buruk, serta kondisi gizi buruk.

3.4.1. Distribusi Kelas Gizi

Sebelum melakukan proses pembentukan pohon keputusan menggunakan algoritma C4.5, terlebih dahulu dilakukan pengelompokan data berdasarkan kelas target, yaitu status gizi. *Dataset* yang digunakan terdiri dari 386 data balita.

Tabel 3.2. Distribusi Kelas Status Gizi

Status Gizi	Jumlah data
Gizi baik	343
Berisiko Gizi Buruk	26
Gizi Buruk	17
Total	386

Rumus *entropy* yang digunakan adalah:

$$\text{entropy}(s) = - \sum_{i=1}^n p_i \log_2(p_i)$$

Keterangan:

S = Himpunan data;

n = jumlah kelas;

P_i = proporsi jumlah data pada kelas ke- i .

Berdasarkan *dataset* yang terdiri dari 386 data, diperoleh

$$\text{Entropy}(S) = - \left(\frac{343}{386} \log_2 \frac{343}{386} \right) - \left(\frac{26}{386} \log_2 \frac{26}{386} \right) - \left(\frac{17}{386} \log_2 \frac{17}{386} \right)$$

Selanjutnya, masing-masing komponen dihitung sebagai berikut:

1. Gizi baik

$$\frac{343}{386} = 0,88860$$

$$- (0,88860 \times \log_2(0,88860)) = 0.15141$$

2. Berisiko gizi buruk

$$\frac{26}{386} = 0,06736$$

$$- (0,06736 \times \log_2(0,06736)) = 0,26216$$

3. Gizi buruk

$$\frac{17}{386} = 0,04404$$

$$- (0,04404 \times \log_2(0,04404)) = 0,19841$$

Maka nilai entropi total adalah:

$$\text{Entropy}(S) = 0,15141 + 0,26216 + 0,19841$$

$$\text{Entropy}(s) = 0,61197$$

Dengan demikian, nilai entropi total dari 386 data balita adalah sebesar 0,61197. Nilai ini selanjutnya digunakan sebagai nilai awal dalam perhitungan *information gain* untuk masing-masing atribut, yaitu umur, jenis kelamin, berat badan, tinggi badan, dan LILA.

3.4.2. Perhitungan Atribut Umur

Atribut umur merupakan atribut numerik yang digunakan dalam proses klasifikasi status gizi balita. Pada algoritma C4.5, atribut numerik perlu dilakukan proses pemisahan menggunakan *threshold* atau titik pemisah. *Threshold* digunakan untuk membagi data menjadi 2 kelompok berdasarkan nilai atribut.

Penentuan *threshold* dilakukan dengan mencari nilai tengah antara dua nilai umur yang berurutan. Rumus yang digunakan adalah:

$$Threshold = \frac{Nilai_i + Nilai_{i+1}}{2}$$

Setiap kandidat threshold kemudian diuji dengan menghitung nilai *Entropy*, *Information Gain*, *Split Information*, dan *Gain Ratio*. *Threshold* yang menghasilkan nilai *Gain Ratio* tertinggi dipilih sebagai *threshold* terbaik untuk atribut umur.

a. Kandidat threshold atribut umur

Berdasarkan *dataset* balita, terdapat beberapa nilai umur yang dapat digunakan untuk membentuk kandidat *threshold*. Dari kandidat yang diuji, diperoleh beberapa kandidat dengan nilai *gain ratio* tertinggi sebagai berikut.

Tabel 3.3. Kandidat Atribut Umur

NO	Threshold (bulan)	Data < Threshold	Data > Threshold	Gain	Gain Ratio
1	6,5	22	364	0,01002	0,03176
2	25,5	155	231	0,03457	0,03557

3	26,5	161	225	0,03190	0,03255
4	36,5	239	147	0,03196	0,03334
5	37,5	245	141	0,03001	0,03169
6	38,5	249	137	0,02880	0,03069
7	58,5	377	9	0,00561	0,03511

Berdasarkan hasil pengujian kandidat *threshold*, nilai 25,5 bulan menghasilkan *gain ratio* tertinggi, yaitu 0,03557. Meskipun *threshold* 58,5 bulan menghasilkan nilai yang mendekati, yaitu 0,03522, nilai tersebut masih lebih rendah dibandingkan dengan *threshold* 25,5 bulan.

Dengan demikian, *threshold* terbaik untuk atribut umur adalah:

$$\text{Threshold} = 25,5 \text{ bulan}$$

Selanjutnya, *dataset* dibagi menjadi dua kelompok:

$$\text{Umur} < 25,5$$

dan

$$\text{Umur} > 25,5$$

b. Distribusi Data Berdasarkan Threshold

berdasarkan *threshold* 25,5 bulan, diperoleh distribusi data sebagai berikut.

Tabel 3.4 Distribusi Data Berdasarkan Umur

Kelompok Umur	Gizi Baik	Berisiko Gizi Buruk	Gizi Buruk	Total
< 25,5 bulan	132	8	15	155
> 25,5 bulan	211	18	2	231
Total	343	26	17	386

C. Entropy Kelompok Umur < 25,5 bulan

Pada kelompok umur < 25,5 bulan terdapat 155 data, terdiri dari 132 data gizi baik, 8 data berisiko gizi buruk, dan 15 data gizi buruk.

$$\begin{aligned} Entropy(S_1) = & -\left(\frac{132}{155} \log_2 \frac{132}{155}\right) - \left(\frac{8}{155} \log_2 \frac{8}{155}\right) \\ & - \left(\frac{15}{155} \log_2 \frac{15}{155}\right) \end{aligned}$$

Sehingga diperoleh:

$$Entropy(S_1) = 0,74410$$

d. Entropy kelompok umur > 25,5 bulan

Pada kelompok umur > 25,5 bulan terdapat 231 data, terdiri dari 211 data gizi baik, 18 data berisiko gizi buruk, dan 2 data gizi buruk.

$$\begin{aligned} Entropy(S_2) = & -\left(\frac{211}{231} \log_2 \frac{211}{231}\right) - \left(\frac{18}{231} \log_2 \frac{18}{231}\right) \\ & - \left(\frac{2}{231} \log_2 \frac{2}{231}\right) \end{aligned}$$

Sehingga:

$$Entropy(S_2) = 0,46556$$

e. Weighted Entropy Atribut Umur

Weighted Entropy dihitung untuk mengetahui nilai *entropy* setelah *dataset* dibagi berdasarkan *threshold* yang telah ditentukan.

$$W E(Umur) = \left(\frac{155}{386} \times 0,74410\right) + \left(\frac{231}{386} \times 0,46556\right)$$

$$W E(Umur) = 0,57741$$

f. Information Gain Atribut Umur

Information gain dihitung dengan mengurangi *entropy* total dengan *weighted*

entropy. *Entropy* total yang telah diperoleh sebelumnya adalah:

$$Entropy(S) = 0,61197$$

Sehingga:

$$Gain(Umur) = Entropy(S) - W E(Umur)$$

$$Gain(Umur) = 0,61197 - 0,57741$$

$$Gain(Umur) = 0,03457$$

g. Split Information Atribut Umur

Split Information digunakan untuk mengetahui informasi dari pembagian data berdasarkan *threshold* yang digunakan.

$$SplitInfo(Umur) = - \left(\frac{155}{386} \log_2 \frac{155}{386} \right) - \left(\frac{231}{386} \log_2 \frac{231}{386} \right)$$

$$SplitInfo(Umur) = 0,97185$$

h. Gain Ratio Atribut Umur

Gain ratio dihitung dengan membagi nilai *information gain* dengan *split information*

$$GainRatio(Umur) = \frac{Gain(Umur)}{SplitInfo(Umur)}$$

$$GainRatio(Umur) = \frac{0,03457}{0,97185}$$

$$GainRatio(Umur) = 0,03557$$

Dengan demikian, threshold terbaik pada atribut umur adalah 25,5 bulan dengan nilai *gain ratio* sebesar 0,03557. *Threshold* tersebut selanjutnya digunakan dalam proses pembentukan pohon keputusan dan dibandingkan dengan *nilai gain ratio* dari atribut lainnya.

3.4.3. Perhitungan Atribut Jenis Kelamin

Atribut jenis kelamin merupakan atribut kategorikal yang terdiri dari dua

kategori, yaitu laki-laki (L) dan perempuan (P). Berbeda dengan atribut numerik seperti umur, atribut jenis kelamin tidak memerlukan penentuan *threshold*. Perhitungan dilakukan dengan membagi *dataset* berdasarkan masing-masing kategori, kemudian menghitung *entropy*, *weighted entropy*, *information gain*, *split information*, dan *gain ratio*.

a. Distribusi Data Berdasarkan Jenis Kelamin

Berdasarkan *dataset* sebanyak 386 data balita, distribusi status gizi berdasarkan jenis kelamin adalah sebagai berikut.

Tabel 3.5 Distribusi Data Berdasarkan Jenis Kelamin

Jenis Kelamin	Gizi Baik	Berisiko Gizi Buruk	Gizi Buruk	Total
Laki-Laki	185	14	12	211
Perempuan	158	12	5	175
Total	343	26	17	386

b. Entropy Jenis Kelamin Laki-Laki

Pada Kelompok laki-laki terdapat 211 data, yang terdiri dari 185 gizi baik, 14 berisiko gizi buruk, dan 12 gizi buruk.

$$Entropy(S_1) = -\left(\frac{185}{211} \log_2 \frac{185}{211}\right) - \left(\frac{14}{211} \log_2 \frac{14}{211}\right) - \left(\frac{12}{211} \log_2 \frac{12}{211}\right)$$

$$Entropy(S_1) = 0,66125$$

c. Entropy Jenis Kelamin Perempuan

Pada kelompok perempuan terdapat 175 data, yang terdiri dari 158 gizi baik, 12 berisiko gizi buruk, dan 5 gizi buruk.

$$Entropy(S_1) = -\left(\frac{158}{175} \log_2 \frac{158}{175}\right) - \left(\frac{12}{175} \log_2 \frac{12}{175}\right) - \left(\frac{5}{175} \log_2 \frac{5}{175}\right)$$

$$Entropy(S_2) = 0,54477$$

d. Weighted Entropy

Weighted entropy dihitung berdasarkan proporsi masing-masing kategori terhadap keseluruhan *dataset*.

$$W E(JenisKelamin) = -\left(\frac{211}{386} \times 66125\right) + \left(\frac{175}{386} \times 0,54477\right)$$

$$W E(JenisKelamin) = 0,60844$$

e. Information Gain

Dengan *Entropy* total sebesar 0,61197, diperoleh:

$$Gain(JenisKelamin) = 0,61197 - 0,60844$$

$$Gain(JenisKelamin) = 0,00353$$

f. Split Information

$$SplitInfo(JenisKelamin) = -\left(\frac{211}{386} \log_2 \frac{211}{386}\right) - \left(\frac{175}{386} \log_2 \frac{175}{386}\right)$$

$$SplitInfo(JenisKelamin) = 0,99372$$

g. Gain Ratio

$$GainRatio(JenisKelamin) = \frac{0,00353}{0,99372}$$

$$GainRatio(JenisKelamin) = 0,00355$$

Berdasarkan hasil perhitungan tersebut, atribut jenis kelamin menghasilkan nilai *gain ratio* sebesar 0,00355. Nilai tersebut selanjutnya dibandingkan dengan nilai *gain ratio* dari atribut lain untuk menentukan atribut yang paling baik dalam pembentukan pohon keputusan.

3.4.4. Perhitungan Atribut Berat Badan

Atribut berat badan merupakan atribut numerik sehingga proses C4.5 dilakukan dengan menemukan *threshold* untuk membagi data menjadi dua kelompok. *Threshold* yang digunakan adalah nilai yang memberikan hasil *gain ratio* terbaik pada atribut berat badan.

Berdasarkan hasil pengujian terhadap dataset, diperoleh *threshold* terbaik sebesar:

$$\text{Threshold} = 8,55 \text{ Kg}$$

Dengan demikian, data dibagi menjadi dua kelompok:

$$\text{Berat} < 8,55 \text{ Kg}$$

Dan

$$\text{Berat} > 8,55 \text{ Kg}$$

a. Distribusi Data Berdasarkan Threshold Berat Badan

Hasil pembagian 386 data berdasarkan *threshold* 8,55 kg adalah:

Tabel 3.6 Distribusi Data Berdasarkan Berat Badan

Kelompok Berat Badan	Gizi Baik	Berisiko Gizi Buruk	Gizi Buruk	Total
< 8,55 Kg	56	4	16	76
> 8,55 Kg	287	22	1	310
Total	343	26	17	386

b. Entropy Berat Badan < 8,55 Kg

Pada kelompok berat badan < 8,55 Kg terdapat 76 data, terdiri dari 56 gizi baik, 4 berisiko gizi buruk, dan 16 gizi buruk.

$$Entropy(S_1) = - \left(\frac{56}{76} \log_2 \frac{56}{76} \right) - \left(\frac{4}{76} \log_2 \frac{4}{76} \right) - \left(\frac{16}{76} \log_2 \frac{16}{76} \right)$$

$$Entropy(S_1) = 1,02146$$

c. Entropy Berat Badan > 8,55 Kg

Pada kelompok berat badan > 8,55 kg terdapat 310 data, terdiri dari 287 gizi baik, 22 berisiko gizi buruk, dan 1 gizi buruk.

$$Entropy(S_1) = - \left(\frac{287}{310} \log_2 \frac{287}{310} \right) - \left(\frac{22}{310} \log_2 \frac{22}{310} \right) - \left(\frac{1}{310} \log_2 \frac{1}{310} \right)$$

$$Entropy(S_2) = 0,40053$$

d. Weighted Entropy

$$W E(Berat) = \left(\frac{76}{386} \times 1,02146 \right) + \left(\frac{310}{386} \times 0,40053 \right)$$

$$W E(Berat) = 0,52278$$

e. Information Gain

Dengan menggunakan *entropy* total:

$$Entropy(S) = 0,61197$$

Maka:

$$Gain(Berat) = 0,61197 - 0,52278$$

$$Gain(Berat) = 0,08919$$

f. Split Info

$$SplitInfo(Berat) = - \left(\frac{76}{386} \log_2 \frac{76}{386} \right) - \left(\frac{310}{386} \log_2 \frac{310}{386} \right)$$

$$SplitInfo(Berat) = 0,71567$$

g. Gain Ratio

$$GainRatio(Berat) = \frac{0,08919}{0,71567}$$

$$GainRatio(Berat) = 0,12463$$

Berdasarkan hasil perhitungan, atribut berat badan menghasilkan *gain ratio* sebesar 0,12463. Nilai tersebut nantinya akan dibandingkan dengan *gain ratio* dari atribut lainnya untuk menentukan atribut yang memiliki pengaruh paling besar.

3.4.5. Perhitungan Atribut Tinggi Badan

Atribut tinggi badan merupakan atribut numerik yang digunakan dalam proses klasifikasi status gizi balita. Oleh karena itu, diperlukan penentuan *threshold* untuk membagi data menjadi 2 kelompok. *Threshold* ditentukan berdasarkan nilai yang menghasilkan *gain ratio* tertinggi dari kandidat *threshold* yang diuji.

Berdasarkan pengujian terhadap *dataset* 386 data balita, diperoleh *threshold* terbaik untuk atribut tinggi badan sebesar:

$$Threshold = 82,45 \text{ cm}$$

Beberapa kandidat dengan nilai *gain ratio* tertinggi dapat dilihat pada tabel berikut.

Tabel 3.7 Kandidat *Threshold* Atribut Tinggi Badan

No	Threshold (cm)	Data < Threshold	Data > threshold	Gain	Gain Ratio
1	82,5	152	234	0,04755	0,04917
2	85,25	185	201	0,04874	0,04880
3	82,65	155	231	0,04674	0,04810
4	85,80	191	195	0,04714	0,04714

5	81,55	135	251	0,04319	0,04626
---	-------	-----	-----	---------	---------

Berdasarkan hasil pengujian tersebut, *threshold* 82,45 cm menghasilkan nilai *gain ratio* tertinggi, yaitu 0,04917. Oleh karena itu, *threshold* tersebut dipilih sebagai titik pemisah terbaik untuk atribut tinggi badan.

a. Data selanjutnya dibagi menjadi:

Tinggi Badan < 82,45 cm

Dan

Tinggi Badan > 82,45 cm

b. Distribusi Data Berdasarkan Threshold

Hasil pembagian berdasarkan *threshold* 82,45 cm adalah sebagai berikut.

Tabel 3.8 Distribusi Berdasarkan Atribut Tinggi Badan

Kelompok Tinggi Badan	Gizi Baik	Berisiko Gizi Buruk	Gizi Buruk	Total
< 82,45 cm	129	7	16	152
> 82,45 cm	214	19	1	234
Total	343	26	17	386

C. Entropy Tinggi Badan < 82,45 cm

Pada kelompok tinggi badan < 82,45 cm terdapat 152 data, terdiri dari 129 data gizi baik, 7 berisiko gizi buruk, dan 16 data gizi buruk.

$$Entropy(S_1) = - \left(\frac{129}{152} \log_2 \frac{129}{152} \right) - \left(\frac{7}{152} \log_2 \frac{7}{152} \right) - \left(\frac{16}{152} \log_2 \frac{16}{152} \right)$$

$$Entropy(S_1) = 0,74727$$

d. Entropy Tinggi Badan > 82,45 cm

Pada kelompok tinggi badan 82,45 cm terdapat 234 data, terdiri dari 214 data gizi baik, 19 data berisiko gizi buruk, dan 1 data gizi buruk

$$\begin{aligned} Entropy(S_2) &= - \left(\frac{214}{234} \log_2 \frac{214}{234} \right) - \left(\frac{19}{234} \log_2 \frac{19}{234} \right) \\ &\quad - \left(\frac{1}{234} \log_2 \frac{1}{234} \right) \\ Entropy(S_2) &= 0,44564 \end{aligned}$$

e. Weighted Entropy

$$\begin{aligned} W E(tinggi) &= \left(\frac{152}{386} \times 0,74727 \right) + \left(\frac{234}{386} \times 0,44564 \right) \\ W E (Tinggi) &= 0,56442 \end{aligned}$$

f. Information Gain

Dengan *entropy* total:

$$Entropy(S) = 0,61197$$

Maka:

$$\begin{aligned} Gain(Tinggi) &= 0,61197 - 0,56442 \\ gain(Tinggi) &= 0,04755 \end{aligned}$$

g. Split Information

$$\begin{aligned} SplitInfo(Tinggi) &= - \left(\frac{152}{386} \log_2 \frac{152}{386} \right) - \left(\frac{234}{386} \log_2 \frac{234}{386} \right) \\ SplitInfo(Tinggi) &= 0,96720 \end{aligned}$$

h. Gain ratio

$$\begin{aligned} GainRatio(Tinggi) &= \frac{0,04755}{0,96720} \\ gainRatio(Tinggi) &= 0,04917 \end{aligned}$$

Berdasarkan hasil perhitungan, atribut tinggi badan dengan *threshold* 82,45 cm menghasilkan nilai *gain ratio* sebesar 0,04917. Nilai tersebut selanjutnya

dibandingkan dengan *gain ratio* atribut lainnya untuk menentukan atribut yang paling baik dalam pembentukan pohon keputusan.

3.4.6. Perhitungan Atribut LILA

Atribut LILA (Lingkar Lengan Atas) merupakan atribut numerik yang digunakan dalam proses klasifikasi status gizi balita. Karena LILA bersifat numerik, maka pada algoritma C4.5 dilakukan penentuan *threshold* untuk membagi data menjadi dua kelompok.

a. Penentuan Threshold Atribut LILA

Threshold diperoleh dari nilai tengah antara dua nilai LILA yang berurutan.

Berdasarkan hasil perhitungan pada *dataset* balita, diperoleh *threshold* terbaik:

$$\text{Threshold} = 11,95 \text{ cm}$$

Dengan demikian, data dibagi menjadi dua kelompok:

$$LILA < 11,95 \text{ cm}$$

Dan

$$LILA > 11,95 \text{ cm}$$

b. Distribusi Data Berdasarkan Threshold LILA

Hasil pembagian data berdasarkan *threshold* 11,95 cm adalah:

Tabel 3.9 Distribusi Data Berdasarkan atribut LILA

Kelompok LILA	Gizi Baik	Berisiko Gizi Buruk	Gizi Buruk	Total
< 11,95 cm	1	1	16	18
> 11,95 cm	342	25	1	368

Total	343	26	17	386
-------	-----	----	----	-----

c. Entropy LILA < 11,95 cm

Pada kelompok LILA < 11,95 cm terdapat 18 data, terdiri dari 1 data gizi baik, 1 data berisiko gizi buruk, dan 16 data gizi buruk.

$$Entropy(S_1) = - \left(\frac{1}{18} \log_2 \frac{1}{18} \right) - \left(\frac{1}{18} \log_2 \frac{1}{18} \right) - \left(\frac{16}{18} \log_2 \frac{16}{18} \right)$$

$$Entropy(S_1) = 0,61437$$

d. Entropy LILA > 11,95 cm

Pada kelompok LILA > 11,95 cm terdapat 368 data, terdiri dari 342 data gizi baik, 25 data berisiko gizi buruk, dan 1 data gizi buruk.

$$Entropy(S_1) = - \left(\frac{342}{368} \log_2 \frac{342}{368} \right) - \left(\frac{25}{368} \log_2 \frac{25}{368} \right)$$

$$- \left(\frac{1}{368} \log_2 \frac{1}{368} \right)$$

$$Entropy(S_2) = 0,38573$$

e. Weighted Entropy

$$W E(LILA) = - \left(\frac{18}{386} \times 0,83291 \right) + \left(\frac{368}{386} \times 0,39397 \right)$$

$$W E(LILA) = 0,39539$$

f. Information Gain

Dengan Entropy total:

$$Entropy(S) = 0,61197$$

Maka:

$$Gain(LILA) = 0,61197 - 0,39539$$

$$Gain(LILA) = 0,21658$$

g. Split Information

$$SplitInfo(LILA) = - \left(\frac{18}{386} \log_2 \frac{18}{386} \right) - \left(\frac{368}{386} \log_2 \frac{368}{386} \right)$$

$$SplitInfo(LILA) = 0,27547$$

h. Gain Ratio

$$GainRatio(LILA) = \frac{0,21658}{0,27547}$$

$$GainRatio(LILA) = 0,78623$$

Berdasarkan hasil perhitungan tersebut, atribut LILA dengan *Threshold* 11,95 cm menghasilkan nilai *gain ratio* sebesar 0,06208. Nilai tersebut selanjutnya dibandingkan dengan *gain ratio* dari seluruh atribut untuk menentukan atribut yang memiliki nilai *gain ratio* tertinggi sebagai *root node* pada pohon keputusan.

3.4.7. Perbandingan Nilai Gain Ratio

Tabel 3.10 Perbandingan Nilai Gain Ratio

No	Atribut	Gain	Split Information	Gain Ratio
1	Umur	0,03457	0,97185	0,03557
2	Jenis Kelamin	0,00353	0,99372	0,00355
3	Berat Badan	0,08919	0,71567	0,12463
4	Tinggi Badan	0,04755	0,96720	0,04917
5	LILA	0,21658	0,27547	0,78623

3.5. Perancangan Sistem

Perancangan sistem merupakan tahap yang dilakukan untuk menggambarkan bagaimana sistem akan dibangun berdasarkan hasil analisis permasalahan. Pada

tahap ini, sistem dirancang untuk dapat membantu petugas kesehatan dalam mengelola data balita serta melakukan proses klasifikasi risiko gizi buruk menggunakan metode C4.5.

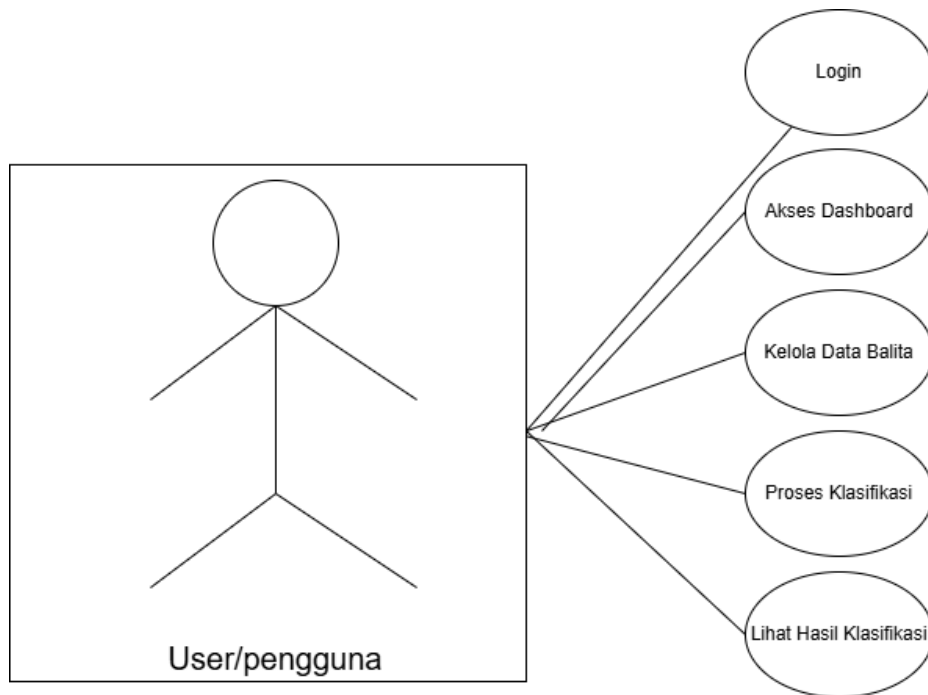
Perancangan sistem dalam penelitian ini menggunakan pendekatan *Unified Modeling Language* (UML) yang bertujuan untuk memodelkan sistem secara visual. Bagan yang diterapkan pada desain tersebut mencakup diagram kasus penggunaan serta diagram aktivitas, yang dimanfaatkan guna memvisualisasikan hubungan pemakai terhadap sistem dan rangkaian tahapan yang berlangsung.

3.5.1. Use Case Diagram

Use case diagram digunakan untuk menggambarkan hubungan antara aktor dengan sistem serta fungsi-fungsi yang dapat dilakukan dalam sistem. Aktor utama dalam sistem ini adalah petugas kesehatan yang bertugas mengelola data balita dan melakukan proses klasifikasi.

Dalam sistem klasifikasi risiko gizi buruk pada balita, aktor dapat melakukan beberapa aktivitas, antara lain:

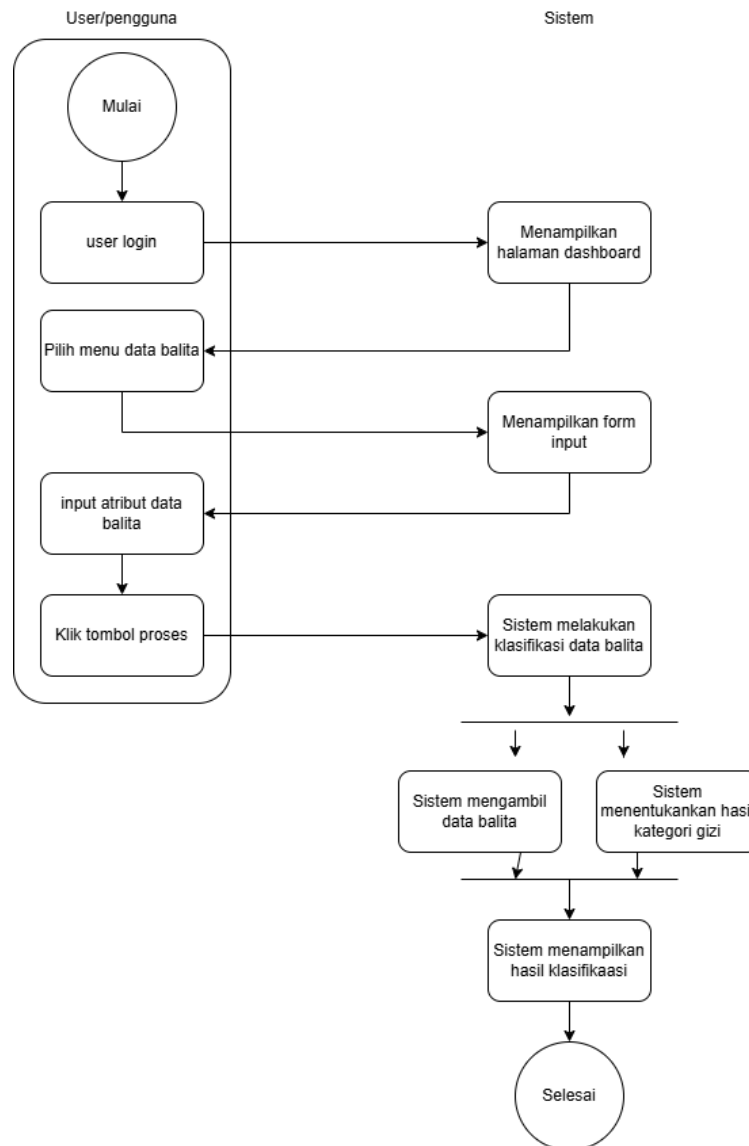
- a. Mengelola data balita (tambah, ubah, hapus data)
- b. Melakukan proses klasifikasi menggunakan metode C4.5
- c. Melihat hasil klasifikasi



Gambar 3.2. *Use Case Diagram*

3.5.2. Activity Diagram

Activity Diagram digunakan untuk menggambarkan alur proses dalam sistem, mulai dari *input* data hingga menghasilkan *output* berupa klasifikasi risiko gizi buruk. Alur proses sistem dimulai dari petugas kesehatan yang memasukkan data balita ke dalam sistem. Data yang telah dimasukkan kemudian disimpan dalam *database*. Berikutnya, sistem akan melaksanakan tahapan pengelompokan dengan menerapkan metode C4.5 berlandaskan variabel yang ada, berupa umur, bobot tubuh, serta tinggi tubuh. Setelah proses klasifikasi selesai, sistem akan menampilkan hasil berupa kategori risiko gizi buruk pada balita. Hasil tersebut dapat digunakan oleh petugas kesehatan sebagai dasar dalam pengambilan keputusan



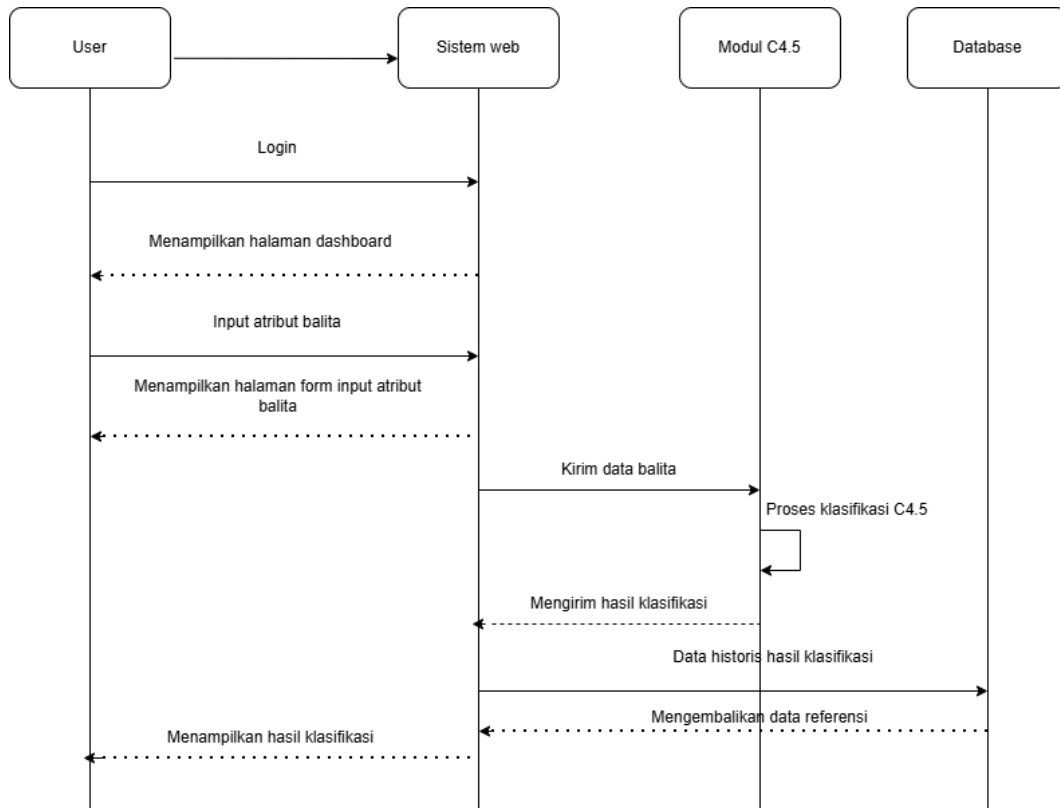
Gambar 3.3. Activity Diagram

Activity diagram yang digunakan pada penelitian ini menggunakan konsep *swimlane*, dengan tujuan agar aktivitas yang dilakukan oleh pengguna sistem dapat dibedakan dengan jelas.

3.5.3. Sequence Diagram

Sequence diagram digunakan untuk menggambarkan interaksi antar objek dalam sistem secara berurutan berdasarkan waktu. Diagram ini menunjukkan bagaimana proses komunikasi terjadi antara aktor (petugas kesehatan) dengan

sistem dalam menjalankan fungsi utama, yaitu proses klasifikasi risiko gizi buruk pada balita menggunakan metode C4.5.



Gambar 3.4. Sequence Diagram

Proses dimulai ketika pengguna melakukan *login* ke dalam sistem melalui antarmuka sistem web dengan memasukkan *username* dan *password*. Sistem kemudian melakukan validasi data *login* ke dalam *database*. Jika data yang di masukan valid, maka sistem web akan menampilkan halaman *dashboard* sebagai tampilan utama aplikasi.

Dari halaman *dashboard*, pengguna memilih menu data balita, kemudian sistem web menampilkan halaman *form input* data balita. Selanjutnya, pengguna *menginput* data balita yang terdiri dari atribut seperti umur, jenis kelamin, berat badan, dan tinggi badan sesuai dengan ketentuan yang telah ditentukan.

Data balita yang telah dimasukkan kemudian dikirim oleh sistem web ke modul klasifikasi untuk dilakukan proses klasifikasi. Modul klasifikasi selanjutnya meneruskan data tersebut ke algoritma C4.5 untuk dilakukan perhitungan.

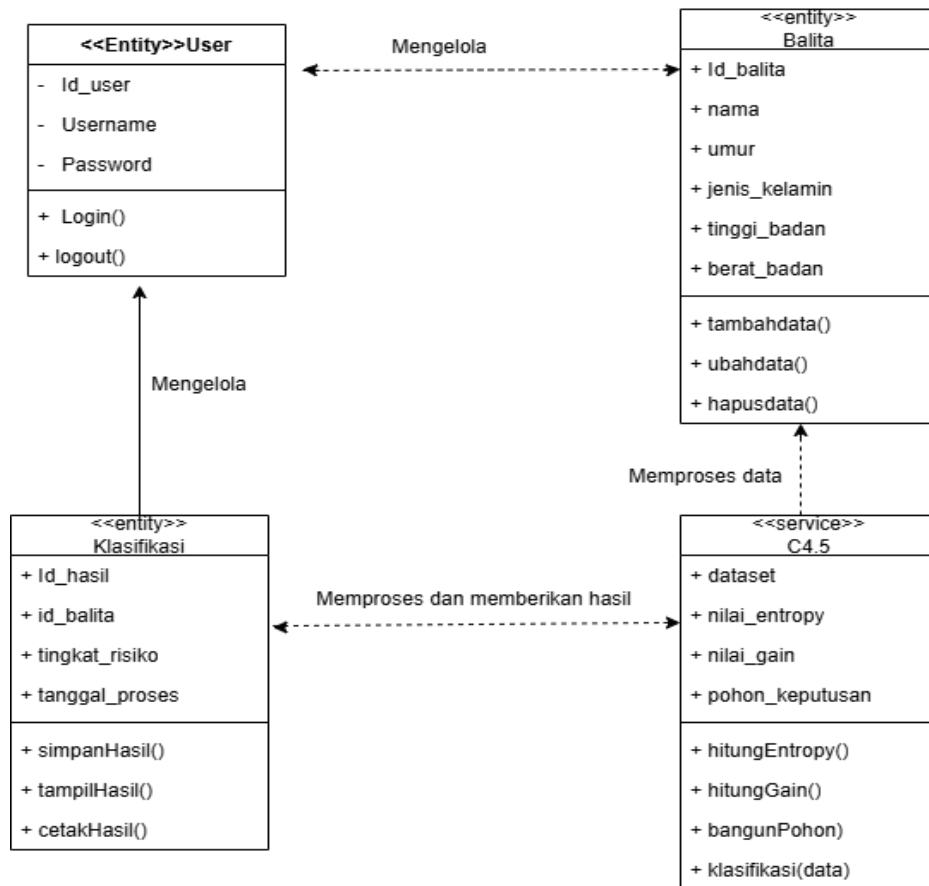
Pada tahap ini, algoritma C4.5 melakukan proses perhitungan nilai *gain entropy* untuk membangun pohon keputusan berdasarkan data yang tersedia. Setelah pohon keputusan terbentuk, sistem akan menentukan hasil klasifikasi berupa tingkat risiko gizi buruk pada balita, seperti risiko rendah, sedang, atau tinggi.

Setelah proses klasifikasi selesai, modul C4.5 mengembalikan hasil klasifikasi ke modul klasifikasi, kemudian diteruskan kembali ke sistem *web*. Selanjutnya menyimpan data hasil klasifikasi ke dalam *database* sebagai data riwayat.

Setelah seluruh proses selesai, sistem web menampilkan hasil klasifikasi kepada pengguna yang meliputi informasi data balita serta hasil klasifikasi tingkat risiko gizi buruk. Dengan demikian, pengguna dapat mengetahui kondisi status gizi balita berdasarkan hasil analisis sistem. Hasil yang ditampilkan diharapkan dapat membantu tenaga kesehatan dalam melakukan deteksi dini terhadap risiko gizi buruk pada balita.

3.5.4. Class Diagram

Class diagram digunakan untuk menggambarkan struktur sistem secara statis yang terdiri dari kelas, atribut, metode, serta hubungan antar kelas dalam sistem. Diagram ini memberikan gambaran mengenai bagaimana data dikelola dan bagaimana interaksi antar komponen dalam sistem klasifikasi risiko gizi buruk pada balita.



Gambar 3.5. Class Diagram

1. *Class User (Entity)*

Kelas ini berfungsi sebagai pengelola akses sistem dan pengendali proses awal dalam penggunaan aplikasi, seperti autentikasi serta pengelolaan data balita dan proses klasifikasi.

balita dan proses klasifikasi.

Atribut pada *class user*:

- 1) *id_user* : identitas unik pengguna
- 2) *username* : nama pengguna yang digunakan untuk *login*
- 3) *Password* : kata sandi pengguna

Method pada *class user*:

- 1) *Login()* : melakukan proses autentikasi pengguna

2) *Logout()* : mengakhiri sesi pengguna

2. *Class* Balita (*Entity*)

Class balita merupakan kelas *entity* yang menyimpan dan mengelola data balita sebagai objek utama dalam penelitian.

Atribut pada *class* balita:

- 1) *id_balita* : identitas unik balita
- 2) *nama* : nama balita
- 3) *umur* : usia balita
- 4) *jenis_kelamin* : jenis kelamin balita
- 5) *tinggi_badan* : tinggi badan balita
- 6) *berat_badan* : berat badan balita

Atribut *berat_badan* dan *tinggi_badan* digunakan sebagai parameter utama dalam proses klasifikasi untuk menentukan kondisi status gizi balita.

Method pada *class* balita:

- 1) *tambahData()* : menambahkan data balita
- 2) *ubahData()* : memperbarui data balita
- 3) *hapusData()* : menghapus data balita

3. *Class* algoritma C4.5 (*Service*)

Kelas ini bertanggung jawab untuk melakukan proses perhitungan *entropy*, *information gain*, pembangunan pohon keputusan, serta menentukan tingkat risiko gizi buruk pada balita.

Atribut pada *class* algoritma C4.5:

- 1) *Dataset* : Kumpulan data balita yang digunakan dalam

proses klasifikasi

- 2) Nilai_entropy : nilai *entropy* hasil perhitungan
- 3) Nilai_gain : nilai *information gain* setiap atribut
- 4) Pohon_keputusan : struktur pohon keputusan hasil algoritma

C4.5

Method pada *class* algoritma C4.5:

- 1) hitungEntropy() : menghitung nilai *entropy*
- 2) hitungGain() : menghitung nilai *information gain*
- 3) bangunPohon() : membangun pohon keputusan
- 4) klasifikasi(data) : menentukan tingkat risiko gizi buruk
berdasarkan data input

Class ini menggunakan atribut seperti umur, jenis kelamin, berat badan, dan tinggi badan sebagai atribut input, sedangkan tingkat risiko gizi buruk digunakan sebagai target klasifikasi

4. *Class* Hasil Klasifikasi (*Entity*)

Class hasil klasifikasi digunakan untuk menyimpan dan menampilkan hasil akhir dari proses klasifikasi risiko gizi buruk yang dilakukan oleh algoritma C4.5.

Atribut pada *class* hasil klasifikasi:

- 1) Id_hasil : identitas unik hasil klasifikasi
- 2) Id_balita : referensi ke data balita yang diklasifikasikan
- 3) Tingkat_risiko : hasil klasifikasi (rendah, sedang, tinggi)
- 4) Tanggal_proses : waktu proses klasifikasi dilakukan

Atribut *tingkat_risiko* merupakan output utama dari sistem yang digunakan untuk mengetahui kondisi gizi balita.

Method pada class hasil klasifikasi:

- 1) *simpanHasil()* : menyimpan hasil klasifikasi ke basis data
- 2) *tampilHasil()* : menampilkan hasil klasifikasi kepada pengguna
- 3) *cetakLaporan()* : mencetak atau mengekspor laporan hasil klasifikasi

3.5.5. Rancangan Antarmuka Aplikasi

Desain antarmuka (User Interface) adalah representasi visualisasi aplikasi yang bakal dimanfaatkan oleh pemakai untuk berkomunikasi terhadap perangkat lunak. Desain tersebut dimaksudkan guna menyediakan kenyamanan pemakaian (*user friendly*) dan menjamin bahwasanya seluruh komponen aplikasi bisa dijangkau secara optimal oleh pemakai.

Pada sistem klasifikasi risiko gizi buruk pada balita, rancangan antarmuka terdiri dari beberapa halaman utama, yaitu halaman *login*, *dashboard*, *input data* balita, serta halaman hasil klasifikasi.

1. Halaman Login

Halaman *login* digunakan oleh pengguna untuk masuk ke dalam sistem dengan memasukkan *username* dan *password*. Halaman *login* merupakan halaman awal yang ditampilkan ketika pengguna mengakses sistem. Pada halaman ini, pengguna diminta untuk memasukkan *username* dan *password*. Sistem kemudian melakukan validasi terhadap data yang dimasukkan. Jika data valid, pengguna akan diarahkan ke halaman *dashboard*.

Logo Puskesmas Gumer

Silahkan Login

Username

Password

Login

Gambar 3.6. Halaman *Login*

2. Halaman Dashboard

Laman *dashboard* adalah laman pokok sesudah pemakai sukses masuk. Laman *dashboard* menyajikan pilihan pokok yang bisa dibuka oleh pemakai, meliputi pengaturan data anak usia bawah lima tahun serta tahapan pengelompokan. Laman tersebut berperan selaku pusat pengarahan pada aplikasi.

Logo Puskesmas gumer Dashboard Pengguna

Selamat Datang Di Web Puskesmas

Dashboard

Data balita

Klasifikasi

Logout

Data Balita Lihat Data

Hasil Klasifikasi Lihat hasil

Gambar 3.7. Halaman *Dashboard*

3. Halaman *Input* Data Balita

Halaman ini digunakan untuk memasukkan data balita yang akan diproses dalam sistem. Pada halaman ini, pengguna dapat memasukkan data balita sesuai dengan atribut yang dibutuhkan dalam proses klasifikasi. Data yang telah *diinput* akan disimpan ke dalam *database* dan digunakan sebagai *input* dalam proses klasifikasi menggunakan algoritma C4.5.

NO	Nama Balita	Umur	JK	BB	TB
1	Balita A	3 Thn	LK	15.6 Kg	88,9 Cm
2	Balita B	4 Thn	P	17,2 Kg	94,1 Cm
3	Balita C	2 Thn	P	10 Kg	77,6 Cm

Gambar 3.8. Halaman Data Balita

4. Halaman Klasifikasi

Halaman klasifikasi memuat proses klasifikasi dan hasil klasifikasi dalam satu tampilan terpadu untuk meningkatkan efisiensi dan kemudahan pengguna sistem. Pada halaman ini, pengguna dapat langsung menjalankan proses klasifikasi data balita dengan menekan tombol klasifikasi. Sistem kemudian menampilkan status proses, dan setelah proses selesai, hasil klasifikasi akan ditampilkan secara langsung pada halaman yang sama. Hasil tersebut meliputi informasi data balita serta tingkat risiko gizi buruk yang dihasilkan oleh algoritma C4.5. Dengan penggabungan ini, pengguna tidak perlu berpindah halaman untuk melihat

hasil, sehingga meningkatkan efektivitas dan kenyamanan dalam penggunaan sistem.

Sistem Klasifikasi Gizi

Puskesmas Gumer

Data Balita

Nama

Umur

Berat badan

Kg

Tinggi Badan

Cm

Jenis Kelamin

☐ Laki-laki

☐ Perempuan

Proses

reset

Hasil Klasifikasi

Status Gizi

Nilai Entropy:

Nilai Gain:

Berdasarkan perhitungan algoritma C4.5, data balita termasuk dalam kategori

Simpan Hasil

Cetak

Riwayat Klasifikasi

No	Nama Balita	Umur	Jenis Kelamin	Berat Badan	Tinggi Badan	Status	Tanggal
1	Balita A	3 Thn	Laki-laki	15,6 Kg	88,9 Cm	Normal	18 Feburari 2026
2	Balita B	4 Thn	Perempuan	17,2 Kg	94,1 Cm	Normal	19 Februari 2026
3	Balita C	2 Thn	Perempuan	10 Kg	77,6 Cm	Berisiko	20 Februari 2026

Gambar 3.9. Halaman klasifikasi

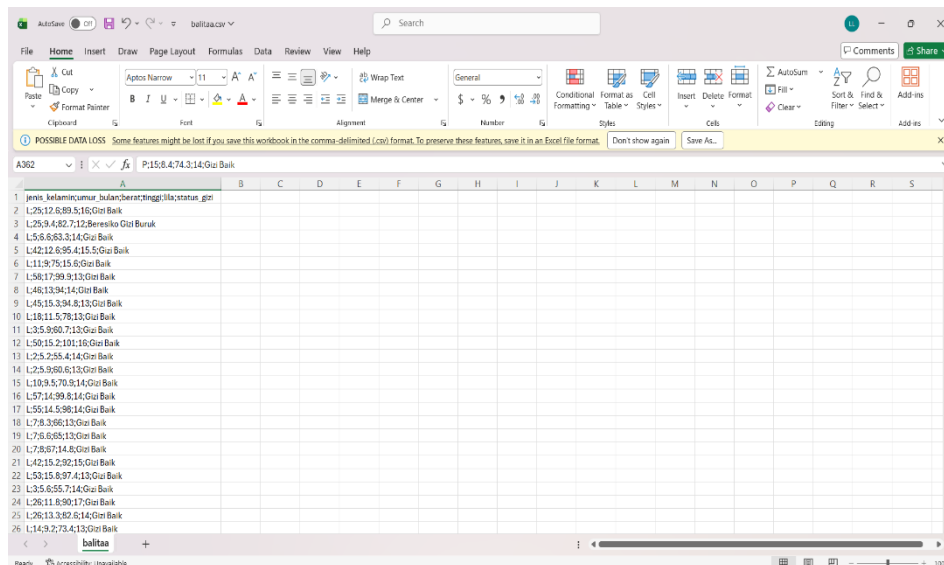
BAB IV

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Implementasi Model dan Dataset

4.1.1 Implementasi Dataset Properti

Dataset yang digunakan dalam penelitian ini merupakan data balita yang diperoleh dari UPTD Puskesmas Gunung Meriah dan digunakan sebagai data dalam proses klasifikasi risiko gizi buruk menggunakan algoritma C4.5. *Dataset* terdiri dari atribut umur, jenis kelamin, berat badan, tinggi badan, dan LILA. Data tersebut kemudian disusun dan diperiksa sebelum digunakan dalam proses klasifikasi. *Dataset* yang telah disiapkan selanjutnya digunakan sebagai sumber data untuk proses pengolahan dan pembentukan model klasifikasi.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	jenis_kelamin:umur_tinggi_badan:LILA:status_gizi																		
2	L:25;12.6;89.5;16;Gizi Baik																		
3	L:25;9.4;82.7;12;Berisiko Gizi Buruk																		
4	L:58;6.83;3.14;Gizi Baik																		
5	L:42;12.6;95.4;15.5;Gizi Baik																		
6	L:11;9.75;15.8;Gizi Baik																		
7	L:58;17.99;9.13;Gizi Baik																		
8	L:48;13.94;14;Gizi Baik																		
9	L:45;15.3;94.8;13;Gizi Baik																		
10	L:18;11.5;78;13;Gizi Baik																		
11	L:3.5;9.60;7.13;Gizi Baik																		
12	L:50;15.2;101.18;Gizi Baik																		
13	L:2.5;2.55;4.14;Gizi Baik																		
14	L:2.5;3.60;6.13;Gizi Baik																		
15	L:10;9.5;70.9;14;Gizi Baik																		
16	L:57;14;99.8;14;Gizi Baik																		
17	L:55;14.5;98.14;Gizi Baik																		
18	L:7.9;3.60;13;Gizi Baik																		
19	L:7.5;6.65;13;Gizi Baik																		
20	L:7.8;67;14.8;Gizi Baik																		
21	L:42;15.2;92;15;Gizi Baik																		
22	L:53;15.8;97.4;13;Gizi Baik																		
23	L:3.5;6.55;7.14;Gizi Baik																		
24	L:26;11.6;90;17;Gizi Baik																		
25	L:26;13.3;92.6;14;Gizi Baik																		
26	L:14;9.2;73.4;13;Gizi Baik																		

Gambar 4.1 Dataset Balita

Gambar 4.1 menunjukkan *dataset* balita yang digunakan dalam penelitian sebagai data masukan untuk proses klasifikasi. *Dataset* tersebut memuat beberapa atribut yang digunakan dalam menentukan kondisi gizi balita, yaitu umur, jenis kelamin, berat badan, tinggi badan, dan LILA. Setiap data balita memiliki nilai

pada masing-masing atribut yang selanjutnya akan diproses menggunakan algoritma C4.5 untuk menentukan kategori klasifikasi, yaitu gizi baik, berisiko gizi buruk, atau gizi buruk. Setelah *dataset* selesai disusun, data kemudian dimasukkan ke dalam *database* MySQL agar dapat dikelola dan digunakan oleh sistem. *Database* berfungsi sebagai tempat penyimpanan data balita yang nantinya digunakan sebagai sumber data dalam proses klasifikasi.

id_balita	jenis_kelamin	umur_bulan	berat	tinggi	ita	status_gizi	kode_balita
2	L	25	12.6	88.5	16	Gizi Baik	BLT002
3	L	25	9.4	82.7	12	Berisiko Gizi Buruk	BLT003
4	L	5	6.6	83.3	14	Gizi Baik	BLT004
5	L	42	12.6	95.4	15.5	Gizi Baik	BLT005
6	L	11	9	75	15.6	Gizi Baik	BLT006
7	L	58	17	99.9	13	Gizi Baik	BLT007
8	L	46	13	94	14	Gizi Baik	BLT008
9	L	45	15.3	94.8	13	Gizi Baik	BLT009
10	L	18	11.5	78	13	Gizi Baik	BLT010
11	L	3	5.9	60.7	13	Gizi Baik	BLT011
12	L	50	15.2	101	16	Gizi Baik	BLT012
13	L	2	5.2	55.4	14	Gizi Baik	BLT013
14	L	2	5.9	60.6	13	Gizi Baik	BLT014
15	L	10	9.5	70.9	14	Gizi Baik	BLT015
16	L	57	14	99.8	14	Gizi Baik	BLT016

Gambar 4.2 Dataset balita pada Mysql

Selanjutnya sistem melakukan koneksi ke *database* MySQL dan mengambil data dari tabel balita menggunakan bahasa pemrograman Python. Proses ini dilakukan agar data yang disimpan di *database* dapat dibaca dan digunakan pada tahap *preprocessing* dan pembentukan model klasifikasi menggunakan algoritma *Decision Tree* C4.5.

```

11 # KONEKSI DATABASE
12 conn = pymysql.connect(
13     host="localhost",
14     user="root",
15     password="",
16     database="gizi_balita"
17 )
18
19 # AMBIL DATA DARI DATABASE
20 query = """
21 SELECT
22     jenis_kelamin,
23     umur_bulan,
24     berat,
25     tinggi,
26     lila,
27     status_gizi
28 FROM balita
29 """
30
31 df = pd.read_sql(query, conn)

```

Gambar 4.3 implementasi Baca dataset dari Mysql

4.2 Implementasi Prosesing Data

Pada tahap *preprocessing*, data yang telah diambil dari *database* MySQL perlu melalui proses transformasi agar dapat digunakan dalam pembentukan model klasifikasi. Proses ini penting dilakukan karena algoritma *Decision Tree* C4.5 memerlukan *input* data dalam bentuk numerik, sedangkan *dataset* yang digunakan masih terdapat atribut dengan tipe data kategorikal. Proses transformasi data dilakukan menggunakan *library* *sklearn.preprocessing* dari Python, khususnya fungsi *LabelEncoder()*. *LabelEncoder* berfungsi untuk mengubah data kategorikal menjadi representasi numerik. Pada penelitian ini, terdapat dua atribut yang mengalami proses *encoding*, yaitu atribut *jenis_kelamin* dan atribut *status_gizi*.

```

# ENCODING DATA

encoder_jk = LabelEncoder()
encoder_status = LabelEncoder()

df["jenis_kelamin"] = encoder_jk.fit_transform(df["jenis_kelamin"])
df["status_gizi"] = encoder_status.fit_transform(df["status_gizi"])

```

Gambar 4.4 Implementasi Preprocessing Data

Gambar 4.3 memperlihatkan proses *encoding* yang dilakukan pada *dataset*. Metode `fit_transform()` digunakan untuk menyesuaikan pola *encoder* terhadap data sekaligus mengubahnya menjadi representasi numerik. Hasil dari proses ini berupa data numerik yang siap digunakan sebagai input pada tahap pelatihan model klasifikasi.

4.3 Implementasi Algoritma Decision Tree C4.5

Penyusunan model pengelompokan dalam riset tersebut menerapkan metode Pohon Keputusan C4.5 yang dioperasikan dengan bahasa komputer Python melalui penggunaan pustaka Scikit-learn. Tahapan tersebut dimaksudkan guna menyusun pola keterkaitan antara ciri-ciri anak usia bawah lima tahun terhadap kondisi gizi yang terdapat. Karakteristik yang digunakan sebagai variabel *input* meliputi jenis kelamin, umur dalam bulan, berat badan, tinggi badan, dan lingkaran lengan atas (LiLA). Sedangkan variabel *output* atau target yang ingin diprediksi adalah status gizi balita yang terbagi menjadi tiga kelas, yaitu Gizi Baik, Berisiko Gizi Buruk, dan Gizi Buruk.

Dataset yang digunakan dalam proses pelatihan model berjumlah [jumlah data] data balita yang telah melalui tahap *preprocessing* sebelumnya. Keseluruhan data tersebut dimanfaatkan untuk melatih model agar mampu mengenali pola dan menghasilkan keputusan klasifikasi yang tepat.

```
# MENENTUKAN FITUR DAN TARGET
x = df[[
    "jenis_kelamin",
    "umur_bulan",
    "berat",
    "tinggi",
    "lila"
]]
y = df["status_gizi"]

# MEMBUAT MODEL C4.5
model = DecisionTreeClassifier(
    criterion="entropy",
    random_state=42
)
```

Gambar 4.5 implementasi Model Decesion Three c4.5

Gambar 4.4, terlihat bahwa proses pembentukan model dilakukan dengan mendefinisikan variabel X sebagai kumpulan fitur dan variabel y sebagai label target. Selanjutnya, objek `DecisionTreeClassifier()` diinisialisasi dengan parameter `criterion="entropy"` yang menjadi ciri khas algoritma C4.5 dalam menghitung pemilihan atribut terbaik. Proses `fit()` kemudian dijalankan untuk melatih model menggunakan data yang telah disiapkan. Hasil dari tahap ini adalah sebuah model pohon keputusan yang siap digunakan untuk mengklasifikasikan data balita baru.

Rule Decision Tree C4.5

```
|--- lila <= 11.95
|   |--- berat <= 8.10
|   |   |--- class: Gizi Buruk
|   |--- berat > 8.10
|   |   |--- class: Gizi Baik
|--- lila > 11.95
|   |--- berat <= 9.15
|   |   |--- class: Beresiko Gizi Buruk
|   |--- berat > 9.15
|   |   |--- class: Gizi Baik
```

Gambar 4.6 Rule Decesion Tree

Berdasarkan Gambar 4.6, *rule decision tree* yang dihasilkan menunjukkan bahwa atribut LILA (Lingkar Lengan Atas) menjadi *root node* atau simpul utama dalam pengambilan keputusan. Jika nilai LILA kurang dari atau sama dengan 11,95 cm, maka sistem akan melakukan pengecekan terhadap atribut berat badan. Apabila berat badan kurang dari atau sama dengan 8,10 kg, maka balita diklasifikasikan ke dalam status Gizi Buruk. Namun, jika berat badan lebih dari

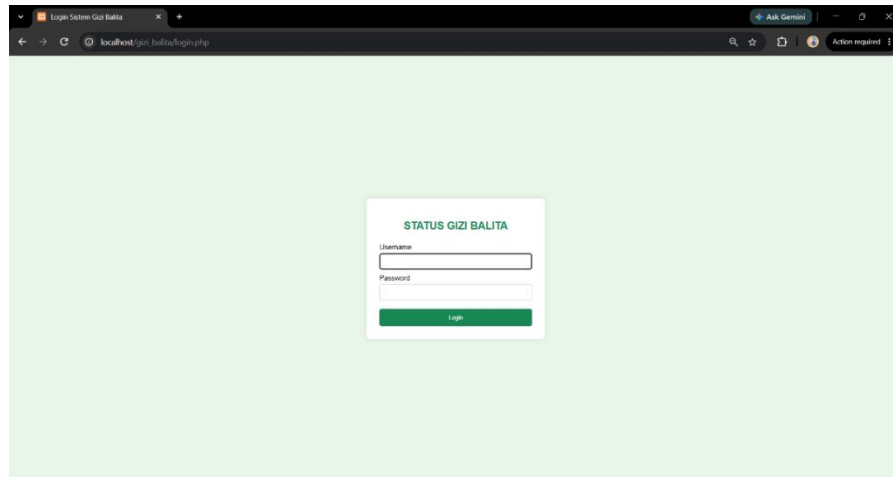
8,10 kg, maka balita diklasifikasikan ke dalam status Gizi Baik. Pada cabang lainnya, jika nilai LILA lebih dari 11,95 cm, sistem kembali melakukan pengecekan terhadap berat badan. Jika berat badan kurang dari atau sama dengan 9,15 kg, maka balita diklasifikasikan ke dalam status Berisiko Gizi Buruk. Sedangkan jika berat badan lebih dari 9,15 kg, maka balita diklasifikasikan ke dalam status Gizi Baik. Dari hasil pembentukan *decision tree* tersebut, dapat diketahui bahwa atribut LILA menjadi faktor paling dominan dalam menentukan status gizi balita, diikuti oleh atribut berat badan sebagai faktor pendukung. Aturan ini kemudian digunakan oleh sistem untuk melakukan klasifikasi secara otomatis terhadap data balita baru yang dimasukkan oleh pengguna.

4.4 Implementasi Antarmuka Sistem

Aplikasi web yang telah dirancang kemudian diimplementasikan dengan menggunakan PHP dan MySQL. Seluruh tampilan disusun sedemikian rupa agar pengguna dapat menjalankan setiap fungsi yang disediakan Terdapat enam halaman utama yang saling terhubung, yaitu halaman masuk, halaman beranda, halaman kelola data balita, halaman proses analisis, halaman tampilan hasil, dan halaman unduh laporan. Masing-masing halaman memiliki peran tersendiri dalam mendukung alur kerja sistem secara keseluruhan

4.4.1. Antarmuka Sistem Admin

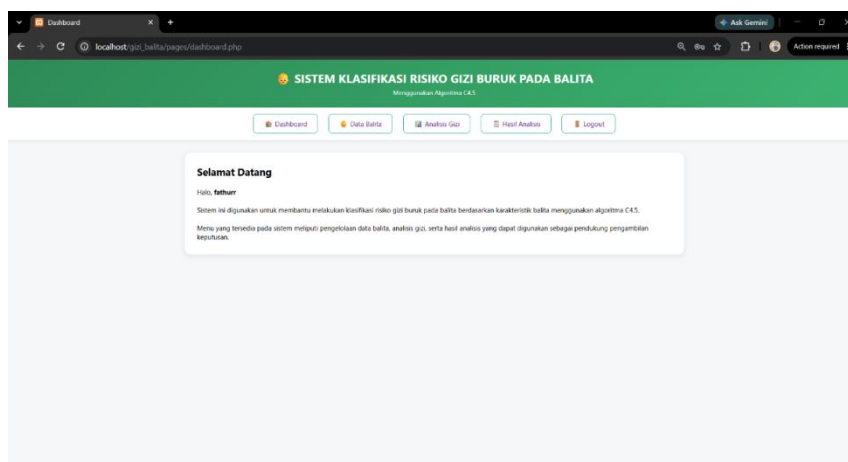
1. Halaman Login Admin



Gamabar 4.7 Halaman Login Admin

Gambar 4.7 Halaman *Login* Admin diminta untuk mengisi *username* dan *password* pada kolom yang telah disediakan. Setelah mengisi data, admin menekan tombol *login* untuk masuk ke sistem. Jika *username* dan *password* yang dimasukkan benar, sistem akan mengarahkan pengguna ke halaman *dashboard*. Namun jika salah, sistem akan menampilkan pesan peringatan dan meminta pengguna mengisi ulang data *login*.

2. Halaman Dashboard Admin

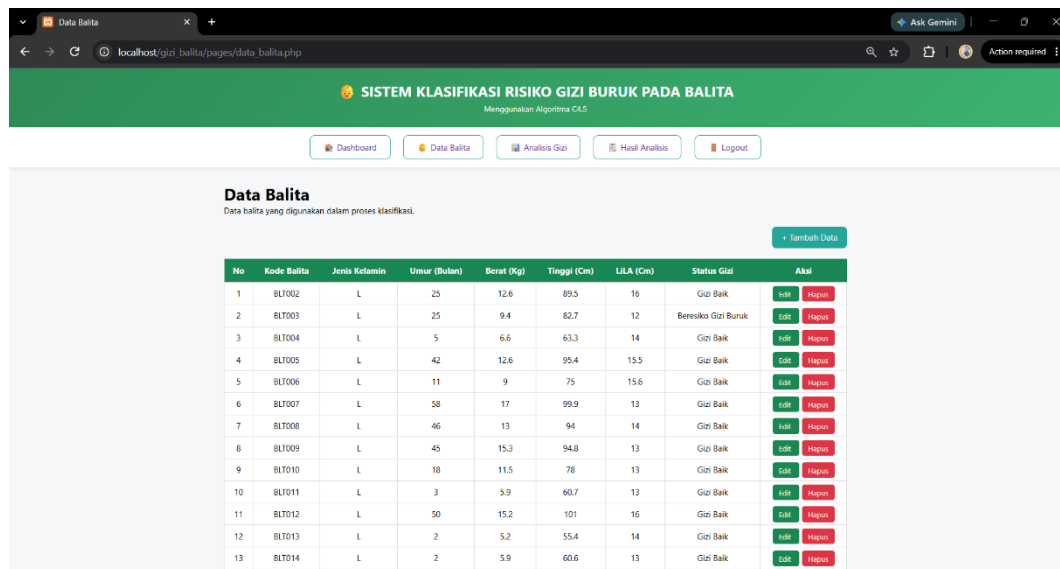


Gambar 4.8 Halaman Dashboard Admin

Gambar 4.8 Halaman *dashboard* Admin merupakan halaman utama yang ditampilkan setelah pengguna berhasil *login* ke dalam sistem. Pada halaman ini,

Admin disambut dengan tampilan yang memuat beberapa menu navigasi untuk mengakses fitur-fitur yang tersedia. Menu-menu tersebut antara lain *Dashboard*, *Data Balita*, *Analisis Gizi*, *Hasil Analisis*, dan *Logout*

3. Halaman Data Balita Admin



No	Kode Balita	Jenis Kelamin	Umur (Bulan)	Berat (Kg)	Tinggi (Cm)	LiLA (Cm)	Status Gizi	Aksi
1	BLT002	L	25	12.6	89.5	16	Gizi Baik	Edit Hapus
2	BLT003	L	25	9.4	82.7	12	Beresiko Gizi Buruk	Edit Hapus
3	BLT004	L	5	6.6	63.3	14	Gizi Baik	Edit Hapus
4	BLT005	L	42	12.6	95.4	15.5	Gizi Baik	Edit Hapus
5	BLT006	L	11	9	75	15.6	Gizi Baik	Edit Hapus
6	BLT007	L	58	17	99.9	13	Gizi Baik	Edit Hapus
7	BLT008	L	46	13	94	14	Gizi Baik	Edit Hapus
8	BLT009	L	45	15.3	94.8	13	Gizi Baik	Edit Hapus
9	BLT010	L	18	11.5	78	13	Gizi Baik	Edit Hapus
10	BLT011	L	3	5.9	60.7	13	Gizi Baik	Edit Hapus
11	BLT012	L	50	15.2	101	16	Gizi Baik	Edit Hapus
12	BLT013	L	2	5.2	55.4	14	Gizi Baik	Edit Hapus
13	BLT014	L	2	5.9	60.6	13	Gizi Baik	Edit Hapus

Gambar 4.9 Halaman Data Balita Admin

Gambar 4.9 Halaman data balita menampilkan seluruh data balita yang tersimpan di dalam *database*. Data yang ditampilkan meliputi kode balita, jenis kelamin, umur dalam bulan, berat badan, tinggi badan, lingkaran lengan atas (LiLA), dan status gizi masing-masing balita. Seluruh data disajikan dalam bentuk tabel yang memudahkan pengguna dalam melihat dan memeriksa setiap informasi yang tersimpan. Pada bagian atas halaman, terdapat tombol "Tambah Data" yang berfungsi untuk menambahkan data balita baru ke dalam sistem. Selain itu, pada setiap baris data juga tersedia tombol "Edit" dan "Hapus" untuk melakukan perubahan atau penghapusan terhadap data yang sudah ada. Halaman ini menjadi pusat pengelolaan data balita yang nantinya akan digunakan sebagai data latih dalam proses klasifikasi menggunakan algoritma C4.5.

4. Halaman Analisis Gizi Admin

Gambar 4.10 Halaman Analisis Gizi Admin

Gambar 4.10 Halaman analisis gizi merupakan halaman yang digunakan untuk melakukan proses klasifikasi status gizi balita. Pada halaman ini, Admin bisa menginput data balita yang akan dianalisis. Data yang perlu diisi meliputi jenis kelamin yang dipilih melalui pilihan Pria atau Wanita, umur dalam satuan bulan, berat badan dalam kilogram, tinggi badan dalam sentimeter, serta lingkaran lengan atas (LiLA) dalam sentimeter. Setelah seluruh data terisi dengan lengkap, Admin dapat menekan tombol “Analisis Gizi” untuk memulai proses klasifikasi. Sistem kemudian akan memproses data yang dimasukkan menggunakan model *Decision Tree* C4.5 yang telah dilatih sebelumnya untuk menentukan status gizi balita. Selain itu, tersedia pula tombol “Kembali” yang berfungsi untuk mengarahkan admin kembali ke halaman *dashboard* jika ingin membatalkan proses analisis

5. Halaman Hasil Analisis Admin

SISTEM KLASIFIKASI RISIKO GIZI BURUK PADA BALITA
Menggunakan Algoritma C4.5

Dashboard Data Balita Analisis Gizi Hasil Analisis Logout

Hasil Analisis Gizi [Cetak Laporan](#)

No	Jenis Kelamin	Umur	Berat	Tinggi	LiLA	Status Gizi	Tanggal Analisis	Aksi
1	L	13	7.5	65	20	Gizi Baik	2026-07-11 14:11:17	Hapus
2	L	13	5	17	20.9	Gizi Baik	2026-07-11 13:43:05	Hapus
3	L	16	7	77	11	Gizi Buruk	2026-06-27 14:04:21	Hapus
4	L	12	5	7	9	Gizi Baik	2026-06-27 13:57:38	Hapus
5	L	24	6.5	65	11.8	Berisiko Gizi Buruk	2026-06-27 13:50:26	Hapus
6	L	3	10	60	60	Gizi Baik	2026-06-27 13:05:22	Hapus
7	L	3	11	75	13	Gizi Baik	2026-06-27 13:04:57	Hapus
8	L	24	6.5	65	11.8	Berisiko Gizi Buruk	2026-06-27 10:52:05	Hapus
9	L	24	7.5	75	11.5	Gizi Buruk	2026-06-27 10:43:01	Hapus
10	P	24	10	82	13.5	Gizi Baik	2026-06-27 10:42:03	Hapus
11	L	24	12.5	86	15.5	Gizi Baik	2026-06-27 10:31:52	Hapus

Gambar 4.11 Halaman Hasil Analisis Admin

Gambar 4.11 Halaman hasil analisis menampilkan seluruh riwayat klasifikasi status gizi balita yang telah dilakukan. Data yang ditampilkan meliputi nomor urut, jenis kelamin, umur, berat badan, tinggi badan, lingkaran lengan atas (LiLA), status gizi hasil klasifikasi, serta tanggal dan waktu pelaksanaan analisis. Seluruh informasi disajikan dalam bentuk tabel yang tersusun rapi sehingga memudahkan Admin dalam melihat dan memeriksa setiap hasil analisis yang telah tersimpan. Pada bagian atas halaman, terdapat tombol "Cetak Laporan" yang berfungsi untuk mencetak atau mengunduh seluruh data hasil analisis dalam bentuk dokumen PDF. Selain itu, pada setiap baris data juga tersedia tombol "Hapus" yang dapat digunakan untuk menghapus riwayat analisis yang sudah tidak diperlukan. Halaman ini memungkinkan Admin untuk memantau perkembangan status gizi balita dari waktu ke waktu serta menjadikannya sebagai bahan evaluasi dalam pengambilan Keputusan

6. Hasil Laporan PDF Admin



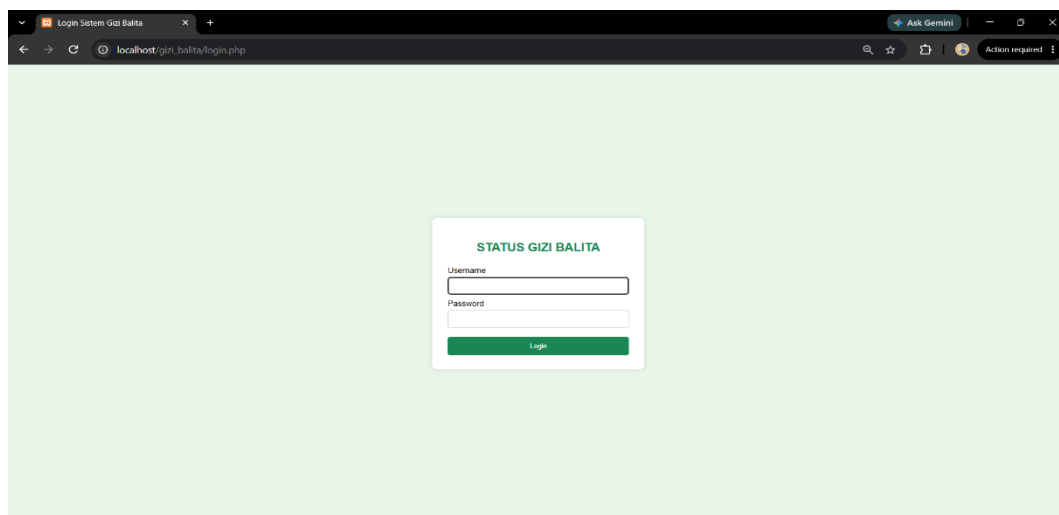
No	Jenis Kelamin	Umur	Berat	Tinggi	LILA	Status Gizi	Tanggal
1	L	13	7.50	65.00	20.00	Gizi Baik	2026-07-11 14:11:17
2	L	13	5.00	17.00	20.90	Gizi Baik	2026-07-11 13:43:05
3	L	16	7.00	77.00	11.00	Gizi Buruk	2026-06-27 14:04:21
4	L	12	5.00	7.00	9.00	Gizi Baik	2026-06-27 13:57:38
5	L	24	6.50	65.00	11.80	Beresko Gizi Buruk	2026-06-27 13:50:26
6	L	3	10.00	60.00	60.00	Gizi Baik	2026-06-27 13:05:22
7	L	3	11.00	75.00	13.00	Gizi Baik	2026-06-27 13:04:57
8	L	24	6.50	65.00	11.80	Beresko Gizi Buruk	2026-06-27 10:52:05
9	L	24	7.50	75.00	11.50	Gizi Buruk	2026-06-27 10:43:01
10	P	24	10.00	82.00	13.50	Gizi Baik	2026-06-27 10:42:03
11	L	24	12.50	96.00	15.50	Gizi Baik	2026-06-27 10:31:52

Gambar 4.12 Hasil Laporan PDF Admin

Gambar 4.12 Halaman ini berfungsi untuk mencetak atau mengunduh seluruh data hasil analisis dalam bentuk dokumen PDF.

4.4.2. Antarmuka Sistem User

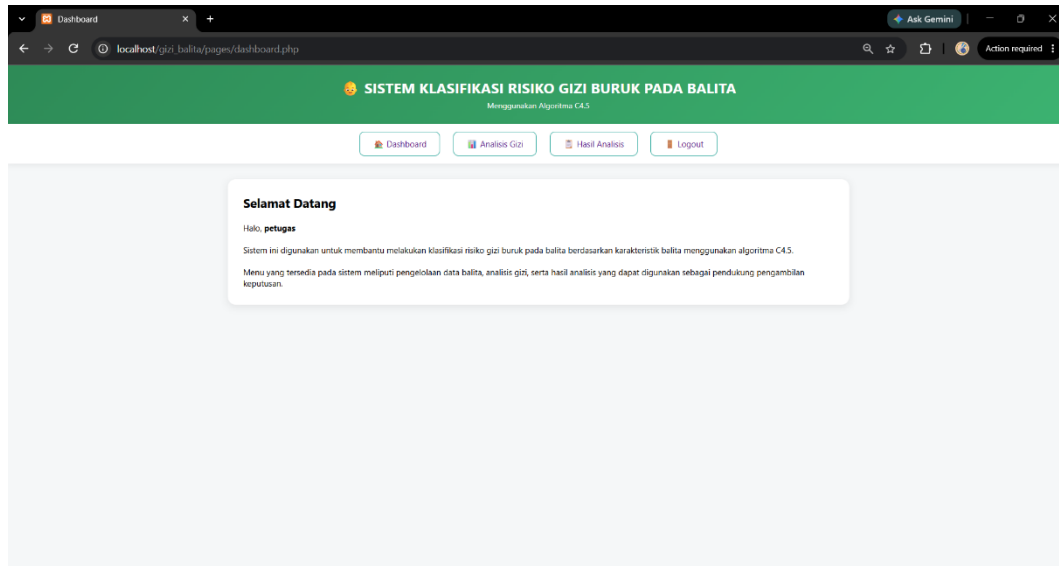
1. Halaman Login User



Gambar 4.13 Halaman *login User*

Gambar 4.11 Laman masuk pengguna adalah laman pertama yang ditampilkan ketika pengguna membuka aplikasi. Di laman tersebut, pengguna diharuskan melengkapi nama pengguna serta kata sandi pada bidang isian yang sudah tersedia.

2. Halaman Dashboard User



Gambar 4.14 Halaman *Dashboard User*

Gambar 4.14 Halaman *Dashboard User*. Setelah berhasil *login*, *user* akan diarahkan ke halaman *dashboard*. Laman tersebut menyajikan pilihan-pilihan menu yang bisa dibuka oleh pengguna berdasarkan kewenangan aksesnya. Pilihan yang disediakan mencakup Dasbor, Pengkajian Gizi, Hasil Pengkajian, serta Keluar.

3. Halaman Analisis Gizi User

Gambar 4.15 Halaman Analisis Gizi *User*

Gambar 4.15 Halaman analisis gizi *user* memiliki fungsi yaitu untuk melakukan proses klasifikasi status gizi balita. ;diminta untuk mengisi formulir

yang berisi data balita yang akan dianalisis, meliputi jenis kelamin, umur, berat badan, tinggi badan, dan LiLA. Setelah seluruh data terisi, *user* menekan tombol "Analisis Gizi" untuk memulai proses klasifikasi.

4. Halaman Hasil Analisis User

The screenshot shows the 'Analisis Gizi Balita' page. At the top, there's a green header with the title 'SISTEM KLASIFIKASI RISIKO GIZI BURUK PADA BALITA' and a subtitle 'Menggunakan Algoritma C4.5'. Below the header are navigation buttons: 'Dashboard', 'Analisis Gizi', 'Hasil Analisis', and 'Logout'. The main content area is titled 'Analisis Gizi Balita' and contains a form with the following fields: 'Jenis Kelamin' (a dropdown menu), 'Umur (Bulan)', 'Berat (Kg)', 'Tinggi (Cm)', and 'LiLA (Cm)'. At the bottom of the form are two buttons: 'Analisis Gizi' and 'Kembali'.

Gambar 4.16 Halaman Hasil Analisis Gizi X

The screenshot shows the 'Hasil Analisis Gizi' page. It features a green header with the same title and subtitle as the previous page. Navigation buttons are also present. The main content area is titled 'Hasil Analisis Gizi' and contains a table with 11 rows of data. A 'Cetak Laporan' button is located at the top right of the table.

No	Jenis Kelamin	Umur	Berat	Tinggi	LiLA	Status Gizi	Tanggal Analisis	Aksi
1	L	13	7.5	65	20	Gizi Baik	2026-07-11 14:11:17	hapus
2	L	13	5	17	20.9	Gizi Baik	2026-07-11 13:43:05	hapus
3	L	16	7	77	11	Gizi Buruk	2026-06-27 14:04:21	hapus
4	L	12	5	7	9	Gizi Baik	2026-06-27 13:57:38	hapus
5	L	24	6.5	65	11.8	Beresiko Gizi Buruk	2026-06-27 13:50:26	hapus
6	L	3	10	60	60	Gizi Baik	2026-06-27 13:05:22	hapus
7	L	3	11	75	13	Gizi Baik	2026-06-27 13:04:57	hapus
8	L	24	6.5	65	11.8	Beresiko Gizi Buruk	2026-06-27 10:52:05	hapus
9	L	24	7.5	75	11.5	Gizi Buruk	2026-06-27 10:43:01	hapus
10	P	24	10	82	13.5	Gizi Baik	2026-06-27 10:42:03	hapus
11	L	24	12.5	86	15.5	Gizi Baik	2026-06-27 10:31:52	hapus

Gambar 4.16 Halaman Analisis Gizi *User* ini menampilkan seluruh riwayat klasifikasi yang telah dilakukan *user*. Data yang ditampilkan meliputi nomor, jenis kelamin, umur, berat badan, tinggi badan, LiLA, status gizi, dan tanggal

analisis. Terdapat tombol "Cetak Laporan" untuk mengunduh hasil dalam bentuk PDF.

5. Hasil Laporan PDF User



LAPORAN HASIL ANALISIS GIZI BALITA
Menggunakan Algoritma C4.5

No	Jenis Kelamin	Umur	Berat	Tinggi	LILA	Status Gizi	Tanggal
1	L	13	7.50	65.00	20.00	Gizi Baik	2026-07-11 14:11:17
2	L	13	5.00	17.00	20.90	Gizi Baik	2026-07-11 13:43:05
3	L	16	7.00	77.00	11.00	Gizi Buruk	2026-06-27 14:04:21
4	L	12	6.00	7.00	8.00	Gizi Baik	2026-06-27 13:57:36
5	L	24	6.50	65.00	11.90	Bersisko Gizi Buruk	2026-06-27 13:50:26
6	L	3	10.00	60.00	60.00	Gizi Baik	2026-06-27 13:09:22
7	L	3	11.00	76.00	13.00	Gizi Baik	2026-06-27 13:04:17
8	L	24	6.50	65.00	11.80	Bersisko Gizi Buruk	2026-06-27 10:62:05
9	L	24	7.50	75.00	11.50	Gizi Buruk	2026-06-27 10:43:01
10	L	24	10.00	82.00	13.50	Gizi Baik	2026-06-27 10:42:03
11	L	24	12.50	86.00	15.50	Gizi Baik	2026-06-27 10:31:52

Gambar 4.17 Hasil Laporan PDF

Gambar 4.17 menunjukkan fitur *download* laporan hasil analisis gizi dalam bentuk PDF. Fitur ini digunakan untuk menyimpan ke dalam bentuk dokumen sehingga dapat digunakan sebagai arsip laporan atau dokumentasi.

4.5 Pengujian Sistem

Pengujian sistem dilakukan untuk mengetahui apakah *website* yang telah dibangun dapat menjalankan setiap fungsi sesuai dengan rancangan yang telah ditentukan. Tahap ini dilakukan dengan mencoba fitur-fitur utama yang tersedia pada sistem dan mengamati hasil yang diberikan setelah pengguna melakukan suatu proses. Pengujian pada penelitian ini dilakukan melalui pengujian fungsional menggunakan Black Box Testing dan pengujian terhadap hasil klasifikasi menggunakan *Confusion Matrix*. Pengujian tersebut digunakan untuk memastikan bahwa sistem dapat digunakan dengan baik serta mengetahui kemampuan algoritma C4.5 dalam melakukan klasifikasi risiko gizi buruk pada balita.

4.5.1. Pengujian Sistem Menggunakan Black Box Testing

Pengujian Black Box dilakukan dengan menguji fungsi sistem bertujuan untuk mengetahui apakah setiap menu dan fitur yang tersedia dapat memberikan keluaran sesuai dengan fungsi yang telah dirancang.

Tabel 4.1 Pengujian BlackBox

No	Fitur yang diuji	Skenario Pengujian	Hasil yang di harapkan	Hasil Pengujian
1	<i>Login</i>	Memasukkan <i>username</i> dan <i>password</i> yang sesuai	Sistem memvalidasi data dan mengarahkan pengguna ke <i>dashboard</i>	Berhasil
2	<i>Dashboard</i>	Membuka halaman <i>dashboard</i> setelah <i>login</i>	Sistem menampilkan halaman utama beserta menu yang tersedia	Berhasil
3	Data Balita	Nambah,edit,hapus data balita	Sistem berhasil menambah, mengedit dan menghapus,	Berhasil
4	Proses Klasifikasi	Menjalankan proses	Sistem	Berhasil

		klasifikasi data balita	memproses data menggunakan algoritma C4.5 dan menghasilkan tingkat risiko	
5	Hasil Klasifikasi	Membuka hasil klasifikasi	Sistem menampilkan hasil klasifikasi risiko gizi buruk	Berhasil
6	Laporan	Menekan tombol <i>cetak/download</i> laporan	Sistem menghasilkan laporan sesuai hasil klasifikasi	Berhasil
7	<i>Logout</i>	Menekan tombol <i>logout</i>	Sistem mengakhiri sesi pengguna dan kembali ke halaman <i>login</i>	Berhasil

4.5.2 Pengujian Hasil Klasifikasi Menggunakan Confusion Matrix

Pengujian hasil klasifikasi dilakukan untuk mengetahui sejauh mana model *Decision Tree* C4.5 mampu mengklasifikasikan status gizi balita secara akurat. Pengujian ini menggunakan metode *confusion matrix* yang membandingkan hasil prediksi sistem dengan data aktual pada *dataset*. *Dataset* yang digunakan sebanyak 386 data balita, kemudian dibagi menjadi data latih (*training*) dan data uji (*testing*) dengan perbandingan 80:20, yaitu 80% data digunakan untuk melatih model dan 20% data digunakan untuk menguji model.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.tree import DecisionTreeClassifier
# Baca data
df = pd.read_csv('balitaa.csv', delimiter=';', encoding='utf-8-sig')

df['lila'] = df['lila'].astype(str).str.replace(',', '.').astype(float)

# MANUAL ENCODE status_gizi
status_map = {
    'Gizi Baik': 0,
    'Gizi Buruk': 1,
    'Beresiko Gizi Buruk': 2
}
df['status_gizi_encoded'] = df['status_gizi'].map(status_map)

df['jenis_kelamin_encoded'] = df['jenis_kelamin'].map({'L': 0, 'P': 1})

X = df[['jenis_kelamin_encoded', 'umur_bulan', 'berat', 'tinggi', 'lila']]
y = df['status_gizi_encoded']
# Split data (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Training model C4.5 (Decision Tree dengan entropy)
model = DecisionTreeClassifier(criterion='entropy', random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

print("="*60)
print("KLASIFIKASI STATUS GIZI BALITA DENGAN C4.5")
print("="*60)
print(f"Akurasi: {accuracy_score(y_test, y_pred) * 100:.2f}%")
print("\nConfusion Matrix:")
print(confusion_matrix(y_test, y_pred))
print("="*60)
```

Gambar 4.18 Script Pengujian Akurasi dan *Confusion Matrix*

Gambar 4.18 menunjukkan *script* pengujian yang digunakan untuk menguji performa model *Decision Tree* C4.5 dalam mengklasifikasikan status gizi balita.

Proses pengujian dimulai dengan membaca *dataset* dari file CSV, kemudian melakukan *preprocessing* data seperti *encoding* pada atribut kategorikal dan pembersihan data LILA. Selanjutnya, *dataset* dibagi menjadi data latih dan data uji dengan perbandingan 80:20 menggunakan fungsi `train_test_split`. Model Decision Tree C4.5 dibangun dengan parameter `criterion='entropy'` yang merupakan ciri khas dari algoritma C4.5. Setelah model dilatih, dilakukan prediksi terhadap data uji dan hasilnya dievaluasi menggunakan *confusion matrix* dan *accuracy score* untuk mengetahui tingkat akurasi model.

```
Akurasi: 87.18%

Confusion Matrix:
[[63  1  3]
 [ 0  5  0]
 [ 5  1  0]]

=====

Detail Confusion Matrix:
              Prediksi
            BGB   GB   GBuruk
Aktual  BGB    63    1     3
        GB     0    5     0
        GBuruk  5    1     0
```

Gambar 4.19 *Confusion Matrix dan Accuracy*

Gambar 4.19 hasil pengujian yang telah dilakukan menggunakan *confusion matrix*, diperoleh tingkat akurasi model sebesar 87,18%. Hasil ini menunjukkan bahwa algoritma C4.5 memiliki kemampuan yang baik dalam mengklasifikasikan status gizi balita ke dalam tiga kategori, yaitu Gizi Baik, Berisiko Gizi Buruk, dan Gizi Buruk.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai klasifikasi risiko gizi buruk pada balita menggunakan metode C4.5 dengan studi kasus di UPTD Puskesmas Gunung Meriah, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil memanfaatkan data antropometri balita yang diperoleh dari UPTD Puskesmas Gunung Meriah untuk mengidentifikasi risiko gizi buruk secara sistematis. Data yang digunakan meliputi atribut umur, jenis kelamin, berat badan, tinggi badan, dan LILA yang diolah menggunakan algoritma C4.5 untuk menghasilkan klasifikasi status gizi balita.
2. Metode C4.5 berhasil diterapkan dalam mengidentifikasi risiko gizi buruk pada balita berdasarkan atribut yang digunakan. Atribut LILA (Lingkar Lengan Atas) merupakan atribut yang paling berpengaruh dalam menentukan klasifikasi status gizi balita dengan nilai *gain ratio* tertinggi sebesar 0,78623, diikuti oleh berat badan (0,12463), tinggi badan (0,04917), umur (0,03557), dan jenis kelamin (0,00355).
3. Berdasarkan hasil pengujian menggunakan *confusion matrix*, model C4.5 menghasilkan tingkat akurasi sebesar 87,18% dalam memprediksi risiko gizi buruk pada balita. Hasil *confusion matrix* menunjukkan bahwa dari 78 data uji, model berhasil mengklasifikasikan 68 data dengan benar dan 10 data dengan kesalahan. Klasifikasi per kelas menunjukkan bahwa kelas Gizi Buruk memiliki akurasi sempurna (100%), kelas Gizi Baik memiliki akurasi 91%, sedangkan kelas Berisiko Gizi Buruk memiliki akurasi 50%. Hal ini disebabkan oleh jumlah data pada kelas Berisiko Gizi Buruk yang relatif sedikit dibandingkan dengan kelas lainnya.

4. Sistem klasifikasi risiko gizi buruk berbasis web yang dibangun mampu membantu tenaga kesehatan dalam mengidentifikasi status gizi balita secara lebih cepat, terstruktur, dan objektif. Sistem dilengkapi dengan fitur pengelolaan data balita, proses analisis menggunakan algoritma C4.5, riwayat hasil klasifikasi, serta laporan dalam bentuk PDF.

5.2 Saran

Adapun saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Penelitian selanjutnya dapat menambahkan atribut lain yang memengaruhi status gizi balita seperti pola makan, riwayat penyakit, tingkat pendidikan orang tua, atau faktor lingkungan agar hasil klasifikasi lebih komprehensif.
2. Sistem dapat dikembangkan dengan membandingkan algoritma C4.5 dengan algoritma klasifikasi lain seperti Naive Bayes, Random Forest, atau K-Nearest Neighbor untuk mengetahui metode yang menghasilkan tingkat akurasi terbaik.
3. Pengembangan sistem ke depan dapat dilakukan dalam bentuk aplikasi berbasis *mobile* agar petugas kesehatan dapat melakukan analisis status gizi secara langsung di lapangan saat kegiatan posyandu.
4. Untuk meningkatkan akurasi model, penelitian selanjutnya disarankan untuk menggunakan *dataset* yang lebih besar dan seimbang antar kelas status gizi, sehingga model yang dihasilkan lebih representatif.

DAFTAR PUSTAKA

- Ariana, R. (2010). *NUTRISI DAN GIZI BURUK*. 4, 60–68.
- Chandra, A. (2023). Pengembangan Aplikasi Berbasis Web Dengan Python Flask Untuk Klasifikasi Data Menggunakan Metode Decision Tree C4.5. *Uii*, 1–84.
- Fahmi, M. N. (2023). Implementasi Mechine Learning menggunakan Python Library : Scikit-Learn (Supervised dan Unsupervised Learning). *Sains Data Jurnal Studi Matematika Dan Teknologi*, 1(2), 87–96.
<https://doi.org/10.52620/sainsdata.v1i2.31>
- Feby, D. (n.d.). Apa Itu Decision Tree di Machine Learning Model? (A. Widya (ed.)). <https://dqlab.id/apa-itu-decision-tree-di-machine-learning-model>
- Hafizan, H., & Putri, A. N. (2020). *Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi*. 1(2), 68–72.
- Islam, H. I. (2022). *Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita Hajar*. 8(July), 116–125.
- Kurniawan, D. (2020). Pengenalan Machine Learning dengan Python. PT Elex Media Komputindo.
- Lestari, D. P. (2022). Upaya Pencegahan Risiko Gizi Buruk pada Balita: Literature Review. *JIUBJ*, 22(1), 532–536.
<https://doi.org/10.33087/jiubj.v22i1.1828>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik*

Informatika, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>

Mardi Yudhi Putra, Nadya Safitri, Nofia Filda Fauziah, Ahmad Safei, & Rayhan Wahyudin Ratu Lolly. (2021). Desain Web Bagi Pemula Menggunakan Framework Bootstrap Pada Smk Taruna Bangsa Bekasi. *Jurnal Buana Pengabdian*, 3(1), 134–148. <https://doi.org/10.36805/jurnalbuanapengabdian.v3i1.1541>

Matondang, T. R., Ramadhan Nasution, Y., Armansyah, & Furqan, M. (2024). Penerapan Algoritma C4.5 Pada Klasifikasi Status Gizi Balita. *Jurnal Fasilkom*, 14(1), 216–225. <https://doi.org/10.37859/jf.v14i1.6941>

Nasution, D. M. (2025). *LITERATURE REVIEW Faktor Risiko Terjadinya Stunting pada Balita DOI : <https://doi.org/10.30596/jih.v6i2.19451>*. 6(2).

Nuraisana, N., Halawa, S. W., & Harun, M. (2024). Implementasi Algoritma C4.5 Dalam Mengklasifikasi Status Gizi Balita Pada Posyandu Desa Sekip Lubuk Pakam. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 23(1), 130. <https://doi.org/10.53513/jis.v23i1.9596>

Patel, B. R., & Rana, K. K. (2014). A Survey on Decision Tree Algorithm For Classification. *Ijedr*, 2(1), 1–5.

Puspitorini, & Sintawati, I. D. (2021). Penerapan Data Minin Untuk Klasifikasi Prediksi Produk Jenis Makanan Kuicng Yang Sesuai Kebutuhan Dengan Alforitma Decision Tree (ID3). *Jurnal AKRAB JUARA*, 6(November), 21–26.

Quinlan, J. R. (1986). “Induction of decision trees.” *Mach Learn* 1, 81–106

(1986). <https://doi.org/10.1007/BF00116251>. *Machine Learning*, 1, 81–106.

Saleh, H. M. . F. R. I. M. (2019). KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN METODE K-NEAREST NEIGHBOR. *KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN METODE K-NEAREST NEIGHBOR*, 4(2), 120–126.

Turban, E., Sharda, R., Delen, D., & King, D. (2011). *Decision Support and Business Intelligence Systems*. Pearson.

Yoshe, M., & Hadikurniawati, W. (2021). *Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita*. 2019.

World Health Organization. (2021). *Malnutrition*.

LAMPIRAN

1. Surat Izin Penelitian



PEMERINTAH KABUPATEN ACEH SINGKIL
DINAS KESEHATAN
UPTD PUSKESMAS GUNUNG MERIAH
DESA LAE BUTAR KEC. GUNUNG MERIAH
ACEH SINGKIL



E-mail: puskesmasgunungmeriah@gmail.com Hotline : 082143539226

Nomor : 445/ 1156 / 2026
Lampiran : -
Perihal : Penelitian

Kepada Yth.
Dekan Fakultas Ilmu Komputer dan Teknologi Informasi
Universitas Muhammadiyah Sumatera Utara
di-

Tempat

Berdasarkan Surat Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara Nomor : 447/IL.3-AU/UMSU-09/F/2026 tanggal 23 Februari 2026, Perihal Izin Penelitian, maka bersama ini kami terangkan bahwa :

Nama : FATHUR RIZQI HASUGIAN
NPM : 2209010209
Judul Penelitian : **KLASIFIKASI KELAYAKAN MAKANAN UNTUK BALITA BERISIKO GIZI BURUK MENGGUNAKAN METODE C4.5 (STUDI KASUS: UPTD PUSKESMAS GUNUNG MERIAH) .**

Telah selesai melakukan Penelitian pada tanggal 24 s.d 25 Februari 2026 di UPTD Puskesmas Gunung Meriah Kabupaten Aceh Singkil.

Demikian surat keterangan ini dibuat untuk dipergunakan sebagaimana mestinya.

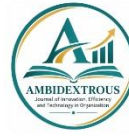
Lae Butar, 25 Januari 2026
Kepala UPTD Puskesmas Gunung Meriah


NPM 2209010209

2. Letter of Acceptance (LoA)



INDO SCHOLARS
Ambidextrous : Journal of Innovation, Efficiency
and Technology in Organization
HP : +62 812-7488-7803 ; E-Mail : admin@indoscholarcenter.com ; E-ISSN : 3031-7002



SURAT KETERANGAN PUBLIKASI ARTIKEL JURNAL

Letter of Acceptance (LoA)

Nomor : /LOA/Ambidextrous/954/Vol5/No3/2026

Kepada YTH.
Fathur Rizqi Hasugian

Terimakasih telah mengirimkan artikel terbaik anda untuk diterbitkan pada **Ambidextrous : Journal of Innovation, Efficiency and Technology in Organization** dengan judul:

"Classification of Malnutrition Risk in Toddlers Using the C4.5 Method (Case Study: UPTD Gunung Meriah Health Center)"

Berdasarkan hasil review dan keputusan tim editor, maka artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Ambidextrous : Journal of Innovation, Efficiency and Technology in Organization** edisi **Volume 5 Nomor 3** bulan **Oktober 2026**.

Demikian surat keterangan ini kami sampaikan untuk dipergunakan sebagaimana mestinya, kami ucapkan terimakasih.

No Reg : LOA20261016054126



Keaslian LOA
Dapat Diperiksa
Dengan Memindai
QR Code
Disamping !

LOA20261016054126

Palembang, 16 October 2026
Editor In Chief



Iizar Daud, Ph.D

Penerbit :
Indo Scholars
Kota Palembang, Sumatera Selatan
+62 812-7488-7803
admin@indoscholarcenter.com



3. Turnitin

SKRIPSI FINAL FATHUR RIZQI SKOM

ORIGINALITY REPORT

28%
SIMILARITY INDEX

25%
INTERNET SOURCES

18%
PUBLICATIONS

14%
STUDENT PAPERS

PRIMARY SOURCES

1	repository.umsu.ac.id Internet Source	2%
2	docplayer.info Internet Source	1%
3	dspace.uji.ac.id Internet Source	1%
4	eprints.upj.ac.id Internet Source	1%
5	eprints.umg.ac.id Internet Source	1%
6	Submitted to Universitas Putera Batam Student Paper	1%
7	Submitted to Universitas Muhammadiyah Palembang Student Paper	1%
8	elibrary.almaata.ac.id Internet Source	1%
9	Submitted to Universitas Islam Riau Student Paper	<1%
10	elib.pnc.ac.id Internet Source	<1%
11	repositori.usu.ac.id Internet Source	<1%
12	123dok.com Internet Source	<1%
13	Fransiskus Febrien, Vivi Aida Fitria. "Sistem Berbasis Logika Fuzzy Mamdani Untuk Klasifikasi Status Gizi Anak", JURNAL FASILKOM, 2026 Publication	<1%
14	Submitted to Sriwijaya University Student Paper	<1%
	eprints.itn.ac.id	