

## **BAB III METODOLOGI PENELITIAN**

### **3.1 Jenis dan Desain Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode penelitian terapan (applied research), metodologi penelitian yang berorientasi pada aplikasi [7]. Pemanfaatan metodologi penelitian kuantitatif memungkinkan peneliti menguji hipotesis secara objektif, menarik generalisasi yang lebih luas tentang populasi, dan memberikan data empiris yang kuat untuk memperkuat kesimpulan mereka [8].

Desain penelitian ini bersifat deskriptif analitik dengan pendekatan cross-sectional, yaitu menganalisis data antropometri balita pada satu periode waktu untuk mengembangkan model prediksi status gizi menggunakan algoritma XGBoost.

### **3.2 Waktu dan Tempat Penelitian**

Penelitian ini dilaksanakan pada periode Januari 2025 hingga Juni 2025. Pengumpulan data dilakukan di RSUD DR. Gondo Suwarno dengan menganalisis rekam medis balita periode 2017-2023. Proses pengembangan sistem dan analisis data dilakukan di Laboratorium Komputer Program Studi Teknik Informatika Universitas Ngudi Waluyo.

### **3.3 Populasi dan Sampel**

#### **3.3.1 Populasi**

Populasi dalam penelitian ini adalah seluruh balita usia 0-59 bulan yang memiliki rekam medis lengkap di RSUD DR. Gondo Suwarno periode 2017-2023.

#### **3.3.2 Sampel**

Penelitian ini menggunakan teknik purposive sampling dalam pengambilan sampel. Menurut Lenaini dkk (2021), purposive sampling merupakan metode non-random sampling di mana peneliti menentukan sampel berdasarkan karakteristik khusus yang sesuai dengan tujuan penelitian [9]. Metode ini memungkinkan untuk memilih sampel secara sengaja sesuai dengan kriteria yang telah ditetapkan untuk melakukan penelitian, sehingga diharapkan dapat menjawab permasalahan penelitian secara efektif [10].

Kriteria Inklusi:

1. Balita usia 0-59 bulan
2. Memiliki data antropometri lengkap (umur, jenis kelamin, tinggi badan, berat badan, lingkar lengan, lingkar kepala)
3. Data tercatat dalam periode 2017-2023

4. Tidak memiliki kelainan kongenital yang mempengaruhi pertumbuhan

Kriteria Eksklusi:

1. Data tidak lengkap atau rusak
2. Outlier ekstrem yang tidak masuk akal secara medis
3. Duplikasi data

Total sampel penelitian sebanyak 5.489 data balita yang memenuhi kriteria.

### **3.4 Pengumpulan Data**

#### **3.4.1 Sumber Data**

Data sekunder berupa rekam medis balita di RSUD DR. Gondo Suwarno periode 2017-2023.

#### **3.4.2 Data**

Dataset penelitian terdiri dari enam atribut utama yang diekstrak dari rekam medis balita:

1. Umur balita (dalam bulan)
2. Jenis kelamin (laki-laki/perempuan)
3. Tinggi badan (dalam cm)
4. Berat badan (dalam kg)
5. Lingkar lengan (dalam cm)
6. Lingkar kepala (dalam cm)

Seluruh pengukuran antropometri dilakukan oleh tenaga kesehatan terlatih mengikuti protokol standar WHO untuk memastikan konsistensi dan akurasi data.

#### **3.4.3 Ethical Clearance**

Penelitian ini telah mendapatkan persetujuan etik dari Komisi Etik Penelitian Universitas Ngudi Waluyo dengan nomor sertifikat 202/KEP/EC/UNW/2025.

### **3.5 Variabel Penelitian**

#### **3.5.1 Variabel Independen**

1. Umur balita (bulan)
2. Jenis kelamin (kategorikal)
3. Tinggi badan (cm)
4. Berat badan (kg)
5. Lingkar lengan (cm)
6. Lingkar kepala (cm)

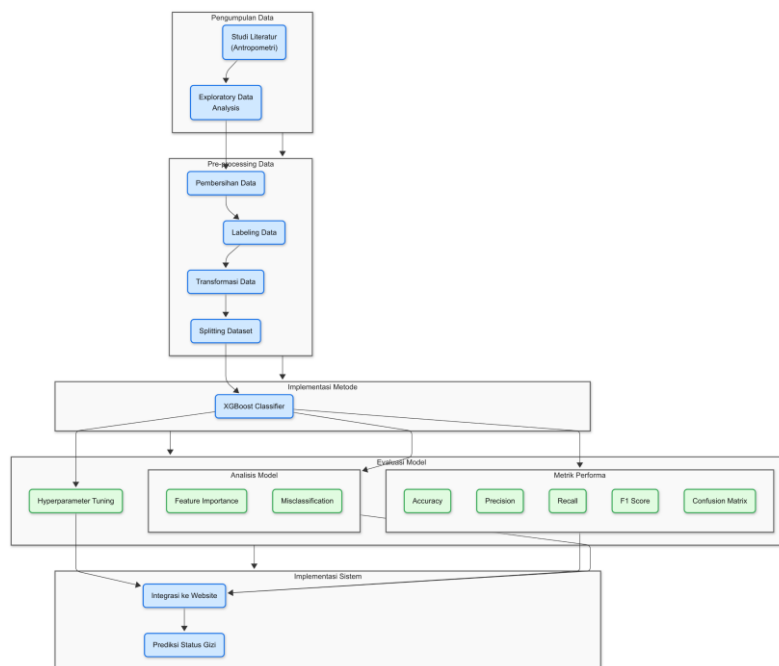
### 3.5.2 Variabel Dependen

Status gizi balita yang dikategorikan menjadi:

1. Gizi Normal (0)
2. Gizi Lebih (1)
3. Gizi Buruk (2)

### 3.6 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis sebagai berikut:



Gambar 1 Tahapan penelitian

1. Pengumpulan Data: Ekstraksi data antropometri balita dari rekam medis
2. Exploratory Data Analysis: Analisis awal untuk memahami karakteristik data
3. Preprocessing Data: Pembersihan dan transformasi data
4. Implementasi Model: Pengembangan model XGBoost
5. Evaluasi Model: Penilaian performa model
6. Implementasi Website: Pengembangan aplikasi web
7. Testing dan Validasi: Pengujian sistem secara keseluruhan

### 3.7 Preprocessing Data

#### 3.7.1 Exploratory Data Analysis

Analisis Data Eksploratif, dalam konteks pendekatan yang disebut Diff In The Loop (DITL) oleh Wang dkk. (2022), merupakan metode interaktif yang menempatkan perbandingan data secara langsung ke dalam siklus

eksplorasi untuk membantu pengguna mengidentifikasi perbedaan penting antar subset data secara iteratif dan mendalam [11]. Proses ini memungkinkan untuk memberikan penyajian ringkasan statistik yang mempermudah memahami informasi pada dataset.

### 3.7.2 Data Cleaning

Cleaning data dalam konteks machine learning yang bersifat eksploratoris, dimana kualitas data berpengaruh sangat signifikan terhadap performa model prediktif [16]. Proses ini memerlukan pendekatan sistematis yang mencakup:

1. Deteksi Missing Values: Identifikasi dan penanganan nilai yang hilang
2. Outlier Detection: Deteksi dan penanganan data pencilan menggunakan metode statistik
3. Duplicate Removal: Eliminasi duplikasi data
4. Data Validation: Validasi konsistensi dan akurasi data

### 3.7.3 Data Labeling

Proses labelling data antropometri balita menggunakan standar baku rujukan Permenkes No. 2 Tahun 2020 [18] diimplementasikan melalui pendekatan nilai z-score untuk tiga indikator utama: TB/U (Tinggi Badan menurut Umur), BB/U (Berat Badan menurut Umur), dan BB/TB (Berat Badan menurut Tinggi Badan).

Ambang batas (Z Score)	TB/U	BB/U	BB/TB
$Z > +3SD$	Tinggi	-	Obesitas
$+2SD < Z \leq +3SD$	-	-	Berisiko Gizi Lebih
$Z > +2SD$	-	Gizi Lebih	-
$-2SD \leq Z \leq +3SD$	Normal	-	-
$-2SD \leq Z \leq +2SD$	-	Gizi Normal	Gizi Normal
$-3SD \leq Z < -2SD$	Pendek	Gizi Kurang	Gizi Kurang
$Z < -3SD$	Sangat Pendek	Gizi Buruk	Gizi Buruk

*Tabel 1 Kategorisasi Status Gizi Berdasarkan Z-Score*

### 3.7.4 Data Transformation

Transformasi adalah langkah yang bertujuan untuk mengubah data masukan menjadi keluaran agar bisa sesuai dengan yang diharapkan [19]. Proses transformasi meliputi:

1. Standardisasi Fitur: Menggunakan StandardScaler untuk menyamakan skala antar variabel
2. Label Encoding: Mengubah data kategorikal menjadi format numerik
3. Feature Engineering: Pembuatan fitur baru jika diperlukan

### 3.7.5 Data Splitting

Pembagian data dilakukan dengan rasio 80:20, yaitu 80% untuk data training dan 20% untuk data testing. Dataset berukuran 100-1.000.000 sampel cocok menggunakan proporsi ini [20]. Parameter stratify digunakan untuk memastikan distribusi kelas yang seimbang pada kedua subset data.

## 3.8 Implementasi Model

### 3.8.1 Algoritma XGBoost

XGBoost classifier dipilih sebagai algoritma utama karena merupakan sistem tree-boosting yang dapat diskalakan dan dirancang khusus untuk meningkatkan performa model machine learning. XGBoost menggabungkan kekuatan teknik gradient boosting dengan regularisasi.

### 3.8.2 Hyperparameter Tuning

Hyperparameter tuning menggunakan metode Grid Search dengan validasi silang sebanyak 3 lipatan (3-fold cross-validation). Proses ini bertujuan menemukan kombinasi parameter yang menghasilkan akurasi terbaik tanpa menyebabkan overfitting.

Parameter	Deskripsi
max_depth	Kedalaman maksimum dari setiap pohon keputusan
learning_rate	Laju pembelajaran dalam boosting untuk menghindari overfitting
min_child_weight	Bobot minimum untuk membentuk cabang baru
subsample	Proporsi data latih yang digunakan pada tiap iterasi boosting
colsample_bytree	Persentase fitur yang digunakan saat membangun pohon

*Tabel 2 Konfigurasi Parameter yang diuji dalam Proses Grid Search*

### 3.9 Evaluasi Model

Evaluasi kinerja model dilakukan melalui pendekatan tiga tingkat yang komprehensif yang dirancang untuk memastikan akurasi dan keandalan dalam mengklasifikasikan status gizi balita.

#### 3.9.1 Metrik Kinerja Dasar

Evaluasi kinerja dasar menggunakan empat metrik utama yang dihitung berdasarkan confusion matrix:

1. Akurasi (Accuracy) mengukur proporsi prediksi yang benar dari total prediksi:
2.  $Accuracy = (TP + TN)/(TP + TN + FP + FN) \dots (1)$
3. Presisi (Precision) mengukur proporsi prediksi positif yang benar:
4.  $Precision = TP/(TP + FP) \dots (2)$
5. Recall (Sensitivitas) mengukur proporsi kasus positif aktual yang berhasil diidentifikasi:
6.  $Recall = TP/(TP + FN) \dots (3)$
7. F1-Score merupakan rata-rata harmonik dari precision dan recall:
8.  $F1-Score = 2 \times (precision \times recall)/(precision + recall) \dots (4)$

#### 3.9.2 Analisis Model Mendalam

##### a) Analisis Kepentingan Fitur (Feature Importance)

Analisis kepentingan fitur merupakan teknik penting dalam proses reduksi dimensi dan memainkan peran krusial dalam keberhasilan implementasi metode data mining dan machine learning. Teknik ini bertujuan untuk memilih subset fitur yang optimal berdasarkan kriteria tertentu [23].

##### b) Misclassification Analysis

Misclassification Analysis dilakukan sebagai salah satu metode evaluasi performa model untuk mengidentifikasi dan memahami jenis kesalahan [24]. Metode ini bertujuan untuk mengungkap pola kesalahan klasifikasi yang paling sering terjadi.

### 3.10 Implementasi Website

#### 3.10.1 Arsitektur Sistem

Aplikasi web dikembangkan dengan arsitektur client-server yang terdiri dari:

1. Frontend: Interface pengguna untuk input data antropometri
2. Backend: Server yang menjalankan model XGBoost
3. Database: Penyimpanan data dan hasil prediksi
4. API: Interface komunikasi antara frontend dan backend

### 3.10.2 Teknologi yang Digunakan

1. Frontend: HTML, CSS, JavaScript
2. Backend: Python dengan framework Flask/Django
3. Machine Learning: Scikit-learn, XGBoost
4. Database: SQLite/MySQL
5. Deployment: Web server (Apache/Nginx)