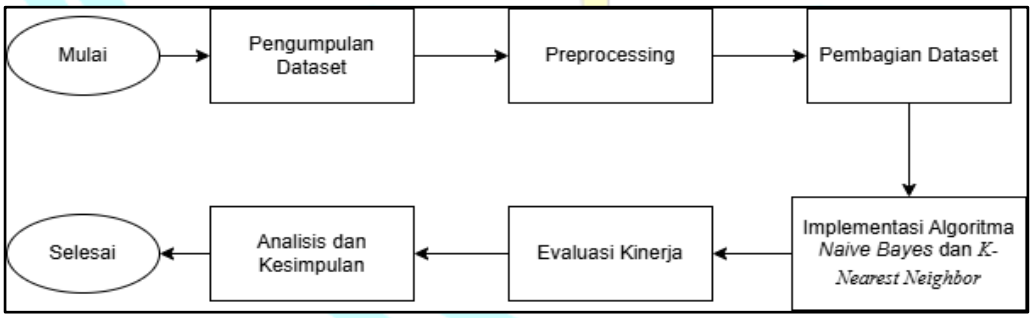
BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Objek dalam penelitian ini yaitu permasalahan terkait status gizi pada anak balita yang ada di Puskesmas Cimahi Selatan. Penelitian ini menggunakan dataset yang terdiri dari 2052 data status gizi anak yang dikumpulkan pada tahun 2024. Dataset tersebut mencakup atribut-atribut penting yaitu, nama, jenis kelamin, usia, berat, tinggi, dan nilai zs bb/tb. Setiap data balita dikelompokan kedalam salah satu dari empat kategori status gizi yaitu gizi buruk, gizi kurang, gizi normal dan gizi lebih. Tujuan penelitian ini untuk membandingkan algoritma *Naive Bayes* dan *K-Nearest Neighbor* untuk klasifikasi status gizi pada anak balita sebagai upaya mendukung menurunkan prevalensi perkembangan gizi pada anak balita yang bisa di terapkan kedalam sebuah sistem teknologi berbasis *data mining*.

3.2 Prosedur Penelitian

Untuk memperoleh hasil yang diharapkan dalam penelitian ini, diperlukan langkah-langkah proses penelitian yang terstruktur. Berikut adalah prosedur yang dilakukan pada proses penelitian ini :



Gambar 3. 1 Prosedur Penelitian

Penjelasan Alur :

1. Pengumpulan Dataset : Data penelitian diperoleh dari Puskesmas Cimahi Selatan, berupa data status gizi anak balita. Data tersebut tersedia dalam format file Excel, yang kemudian diubah menjadi format file CSV untuk mempermudah pengolahan data. Total data yang diperoleh sebanyak 2.052 data anak balita, yang mencakup berbagai informasi terkait status gizi anak. Data ini dikumpulkan dalam rentang waktu bulan Agustus hingga November

2024. Data penelitian yang telah dikumpulkan dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Dataset Penelitian

No	Nama	JK	Usia Saat Ukur	Berat	Tinggi	ZS BB/TB
1	Salsa Nur	P	4 Tahun – 10 Bulan	13	99.8	-1.75
2	Devina	L	4 Tahun – 10 Bulan	16.4	104.1	-0.12
.	.	.	.	.	.	.
.	.	.	.	.	.	.
2051	Meysila	P	2 Tahun – 2 Bulan	10	92.5	-3.17
2052	Raynan Raykkhan	L	2 Tahun – 0 Bulan	8.8	81	-2.53

2. *Preprocessing* : Data diproses lebih lanjut untuk meningkatkan kualitas untuk memastikan data dengan format yang sesuai.

   A. Penambahan atribut Status Gizi : Pada tahap ini, dilakukan penambahan atribut berupa kolom Status Gizi yang berfungsi sebagai label kelas dalam proses klasifikasi. Atribut ini diperoleh berdasarkan nilai *z-score* yang ada pada kolom ZS BB/TB sesuai standar *WHO*. Nilai *z-score* tersebut kemudian dikategorikan kedalam 4 kelas, yaitu Gizi Normal, Gizi Lebih, Gizi Kurang, dan Gizi Buruk. Hasil penambahan kolom dapat dilihat pada Tabel 3.2.

Tabel 3. 2 Penambahan kolom Status Gizi

No	Nama	JK	Usia Saat Ukur	Berat	Tinggi	ZS BB/ TB	Status Gizi
1	Salsa Nur	P	4 Tahun – 10 Bulan	13	99.8	-1.75	Gizi Normal
2	Devina	L	4 Tahun	16.4	104.1	-0.12	Gizi

			– 10 Bulan				Normal
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
2051	Meysila	P	2 Tahun – 2 Bulan	10	92.5	-3.17	Gizi Buruk
2052	Raynan Raykkhan	L	2 Tahun – 0 Bulan	8.8	81	-2.53	Gizi Kurang

B. Pembersihan Data (Data Cleaning) : Menghapus atau memperbaiki data yang tidak sesuai, duplikat, atau hilang. Tabel 3.3 menunjukan contoh penghapusan data duplikat pada dataset yang digunakan. Sedangkan tabel 3.4 menunjukan contoh mengatasi nilai yang hilang atau tidak sesuai.

Tabel 3. 3 Penghapusan Data Duplikat

Data Awal	**Data setelah dilakukan penghapusan duplikat`**
Nama	**Nama**
Salsa Nur	Salsa Nur
Salsa Nur	Devina
Devina	

Tabel 3. 4 Mengatasi Nilai Kosong

Data Awal		**Data yang sudah diatasi terhadap Nilai yang hilang**	
Nama	**Usia Saat Ukur**	**Nama**	**Usia Saat Ukur**
Salsa Nur	4 Tahun – 10 Bulan	Salsa Nur	4 Tahun – 10 Bulan
Devina	4 Tahun – 10 Bulan	Devina	4 Tahun – 10 Bulan
Resa	NaN	Resa	4 Tahun – 10 Bulan

3. Tranformasi Data : Menstandarkan format data, seperti konversi nilai numerik atau kategori menjadi numerik dengan menggunakan *Label Encoding*. Tabel 3.5 menunjukan data sebelum dilakukan *Label Encoding*, sedangkan tabel 3.6 menunjukan data yang sudah dilakukan proses *Label Encoding*.

Tabel 3. 5 Data sebelum dilakukan *Label Encoding*

Nama	**JK**	**Usia Saat Ukur**	**Status Gizi**
Salsa Nur	P	4 Tahun – 10 Bulan	Gizi Normal
Devina	L	4 Tahun – 10 Bulan	Gizi Normal
Resa	P	4 Tahun – 10 Bulan	Gizi Normal

Tabel 3. 6 Data sesudah dilakukan *Label Encoding*

Nama	**Jenis Kelamin**	**Usia Bulan**	**Status Gizi**
Salsa Nur	1	58	3
Devina	0	58	3
Resa	1	58	3

4. Pemilihan Fitur : Tujuan pemilihan fitur adalah untuk membuat data lebih mudah dipahami dan meningkatkan kinerja model. Dengan menggunakan *heatmap*, korelasi antara fitur dan variabel target dianalisis. Fitur-fitur yang kurang berkontribusi atau memiliki hubungan yang terlalu kuat dengan fitur lain diputuskan untuk tidak digunakan untuk menghindari *overfitting* dan mendukung kinerja model yang lebih optimal.
5. *Balancing* Data : Untuk mengatasi masalah jumlah data yang tidak seimbang antar kelas, *balancing* data dilakukan. Jika data tidak seimbang, model lebih sering memprediksi kelas dengan jumlah yang lebih besar, sehingga kelas yang lebih sedikit akan lebih buruk. Metode *over-sampling* minoritas sintetis, atau *SMOTE*, digunakan untuk mengimbangi penelitian ini. Metode ini menghasilkan data baru pada kelas minoritas dengan menambahkan variasi data baru yang sebanding. Oleh karena itu, jumlah data untuk setiap kelas menjadi seimbang.
6. Pembagian Dataset : Setelah dilakukan *balancing* data, total data menjadi 6820

data, kemudian dataset dibagi menjadi dua bagian yaitu data latih untuk melatih model dan data uji untuk mengevaluasi performa model. Pembagian dilakukan dengan proporsi 80% data latih sebanyak 5456 data dan 20% data uji sebanyak 1364 data.

7. Normalisasi Data : adalah proses penyesuaian skala nilai agar semua fitur memiliki rentang yang seragam, sehingga algoritma klasifikasi dapat bekerja lebih optimal. Beberapa metode normalisasi yang umum digunakan adalah *Min-Max Scaling*. Persamaan *Min-Max Scaling* bisa dilihat pada persamaan 7. Tabel 3.7 menunjukan data yang belum di normalisasi, sedangkan Tabel 3.8 menunjukan data yang sudah di normalisasi.

$$\textit{Min-Max Scaling} = X' \frac{X - Xmin}{Xmax - Xmin} \quad (7)$$

Tabel 3. 7 Data sebelum di normalisasi

Nama	Usia Bulan	Berat Badan	Tinggi Badan
Salsa Nur	58	13	99.8
Devina	58	16.4	104.1
Resa	58	18.7	113

Tabel 3. 8 Data yang sudah di normalisasi

Nama	Usia Bulan	Berat Badan	Tinggi Badan
Salsa Nur	0.966667	0.370370	0.742402
Devina	0.966667	0.496296	0.804631
Resa	0.966667	0.581481	0.933430

8. Implementasi Algoritma *Naïve Bayes* dan *K-Nearest Neighbor* : Dalam penelitian ini, dua algoritma klasifikasi dipilih untuk dibandingkan, yaitu *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Kedua algoritma ini dipilih karena memiliki karakteristik yang relevan untuk analisis klasifikasi status gizi.

   A. Langkah – Langkah algoritma *K-Nearest Neighbor* :

      1. Mulai
      2. Siapkan data:

a. Data *training* : digunakan untuk melatih model - X_*train* : fitur (usia saat ukur, jenis kelamin, berat, tinggi dan ZS BB/TB) - Y_*train* : Label (Status gizi : gizi normal,gizi buruk, gizi kurang dan gizi lebih)

b. Data *testing* : digunakan untuk menguji hasil prediksi - X_*test* :Fitur yang digunakan untuk memprediksi label.

3. Tentukan rentang nilai K yang akan diuji (misalnya: 1, 3, 5, 7, dst.)
4. Untuk setiap nilai K : - Lakukan 5-*fold cross-validation* pada data training - Hitung akurasi rata-rata untuk setiap nilai K
5. Pilih Nilai K dengan Akurasi Tertinggi sebagai K Optimal (K=1).
6. Lakukan Klasifikasi pada *Data Testing*: - Untuk setiap data dalam X_*test* (data yang akan diuji):
   a. Hitung jarak antara data uji dengan setiap data dalam X_*train*: - Gunakan rumus jarak *Euclidean*:
   b. Jarak = √((nilai_uji1 - nilai_latih1)^2 + (nilai_uji2 - nilai_latih2)^2 + ...)
   c. Simpan semua jarak yang dihitung.
   d. Urutkan jarak dari yang terkecil ke terbesar.
   e. Pilih 1 tetangga terdekat (karena K=1, hanya data dengan jarak terkecil yang diambil).
7. Tentukan label untuk data uji:
   a. Lihat label dari K tetangga terdekat.
   b. Hitung jumlah kemunculan masing-masing kelas (gizi normal, gizi buruk, gizi kurang dan gizi lebih).
   c. Pilih label yang paling sering muncul (mayoritas) sebagai prediksi.
8. Simpan prediksi hasil untuk setiap data uji.
9. Ulangi langkah 6-8 untuk semua data dalam X_*test*.
10. Selesai

B. Langkah – Langkah algoritma *Naïve Bayes* :

1. Mulai
2. Siapkan data:
   a. Data *training*: - X_*train*: Fitur (usia saat ukur, jenis kelamin, berat,

tinggi, ZS BB/TB) - Y_*train*: Label (status gizi: gizi normal, gizi buruk, gizi kurang, gizi lebih)

b. Data *testing*: - X_*test*: Fitur yang digunakan untuk memprediksi status gizi

3. Hitung probabilitas prior P(C) untuk setiap kelas dalam Y_*train*: P(C) = (jumlah sampel dengan kelas C) / (total jumlah sampel) Hitung rata-rata (μ) dan standar deviasi (σ) untuk setiap fitur dalam setiap kelas C: - μ_C = *mean* (nilai fitur untuk kelas C) - σ_C = std_dev(nilai fitur untuk kelas C) Untuk setiap data dalam X_*test*:
   a. Hitung probabilitas *likelihood* P(X|C) menggunakan distribusi *Gaussian*: P(Xi|C) = (1 / (√(2π) x σ_C)) x exp(- (Xi - μ_C)^2 / (2σ_C^2))
   b. Hitung probabilitas posterior P(C|X) menggunakan Teorema Bayes: P(C|X) ∝ P(C) x P(X1|C) x P(X2|C) x ... x P(Xn|C)
   c. Pilih kelas dengan probabilitas posterior tertinggi sebagai prediksi status gizi.
4. Simpan hasil prediksi untuk setiap data dalam X_*test*.
5. Selesai.

10. Evaluasi Kinerja : Evaluasi dilakukan dengan menggunakan metrik seperti akurasi, *precision*, *recall*, dan F1-*score*. Hasil evaluasi ini akan menunjukkan algoritma mana yang lebih efektif untuk klasifikasi status gizi anak berdasarkan data yang tersedia.
11. Analisis dan Kesimpulan : Hasil evaluasi dibandingkan untuk menentukan keunggulan masing-masing algoritma. Analisis ini digunakan untuk menarik kesimpulan dan memberikan hasil terkait penggunaan algoritma terbaik untuk kasus klasifikasi status gizi anak.