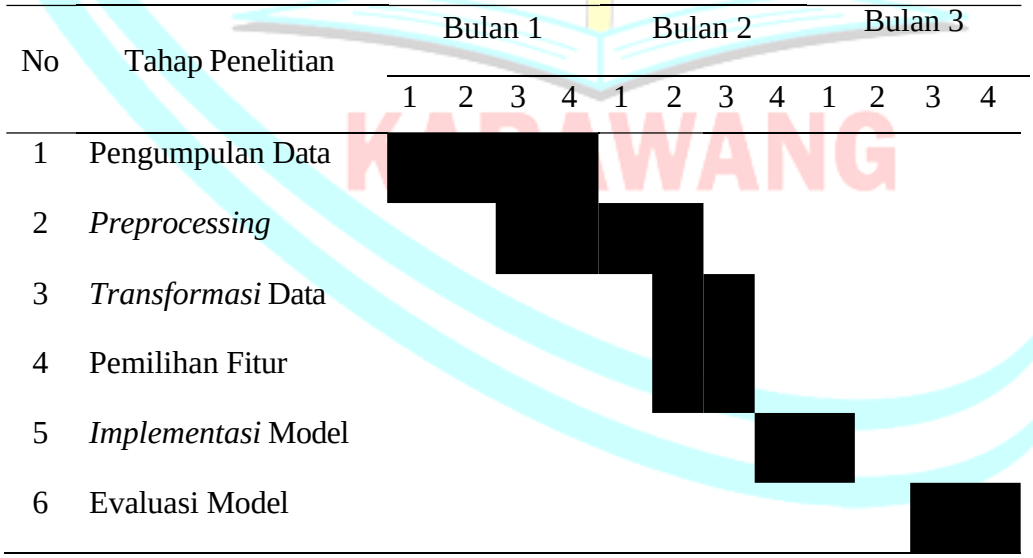
BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian dalam studi ini yaitu kasus *stunting* yang terjadi di Desa Pasirjengkol. Penelitian ini menggunakan dataset yang terdiri dari 505 data pertumbuhan balita yang dikumpulkan pada tahun 2024. Dataset tersebut mencakup atribut-atribut penting, yaitu nama, usia, jenis kelamin, dan tinggi badan. Setiap data balita dikelompokkan kedalam salah satu dari tiga kategori yaitu normal, *stunting*, dan *severely* stunting. Tujuan dari penelitian ini adalah untuk mengklasifikasikan status *stunting* pada balita sebagai upaya mendukung deteksi dini dan penanganan yang lebih efektif.

3.1.1 Lokasi dan Waktu Penelitian

Lokasi penelitian dilakukan di Desa Pasirjengkol, Kecamatan Majalaya, Kabupaten Karawang, Jawa Barat, dengan fokus pada data pemeriksaan balita yang diperoleh dari 12 posyandu di desa pasirjengkol. Waktu penelitian di tunjukkan pada Gambar 3.1.

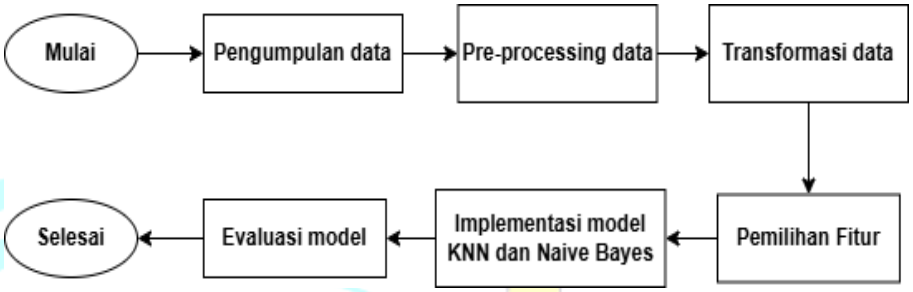


Gambar 3. 1 Waktu Penelitian

3.2 Prosedur Penelitian

Untuk mencapai hasil penelitian yang maksimal, diperlukan pendekatan yang terencana dengan baik dan terstruktur. Proses penelitian ini melibatkan serangkaian

langkah yang dilakukan secara sistematis untuk memastikan setiap tahap dapat dilaksanakan dengan efektif dan saling mendukung. Sebagai gambaran yang lebih jelas mengenai alur penelitian ini, prosedur tersebut disajikan dalam bentuk diagram alur (*flowchart*) yang menggambarkan setiap tahap secara lengkap dan rinci, seperti yang terlihat pada gambar 3.2.



Gambar 3.2 *Flowchart* Prosedur Penelitian

3.3 Pengumpulan data

Pengumpulan data dilakukan di Posyandu yang terletak di Desa Pasirjengkol, dengan total sebanyak 505 data. Data yang dikumpulkan merupakan data pertumbuhan balita dan mencakup empat atribut , yaitu nama, usia (bulan), jenis kelamin, tinggi badan. Proses pengumpulan data dilakukan dengan mencatat langsung dari catatan kesehatan Posyandu yang memuat informasi rutin mengenai pertumbuhan balita. Data penelitian yang dikumpulkan dilihat pada Tabel 3.1.

Tabel 3. 1 Dataset Penelitian

No	Nama	Usia bulan	Jenis Kelamin	Tinggi Badan
1	Mizan	24	Laki - Laki	73
2	Rafa	21	Laki - Laki	78
3	Raizi	18	Laki - Laki	79.5
4	Elvano	20	Laki - Laki	88.8
5	David	19	Laki - Laki	81.3
...	…	…	…	…
505	Ciyra	34	Perempuan	79.7

3.4 *Pre-processing data*

Preprocessing data merupakan tahap penting dalam pengolahan data yang bertujuan untuk meningkatkan kualitas data sehingga dapat digunakan secara optimal dalam pelatihan model klasifikasi. Dalam penelitian ini, dataset yang

digunakan telah diperiksa dan dipastikan tidak mengandung data duplikat maupun nilai kosong, sehingga tahapan pembersihan data (*data cleaning*) seperti penghapusan data duplikat dan penanganan nilai kosong tidak diperlukan.

3.5 *Transformasi* Data

Transformasi data merupakan bagian dari proses *preprocessing* yang bertujuan untuk mempersiapkan data mentah menjadi format yang sesuai untuk digunakan dalam proses pembelajaran mesin. Dalam penelitian ini, *transformasi* data dilakukan melalui beberapa tahapan yaitu menambahkan atribut status, *label encoding* dan *normalisasi* data.

3.5.1 Penambahan atribut status

Pada tahap ini, Dilakukan penambahan atribut baru berupa kolom Status yang berfungsi sebagai label kelas dalam proses klasifikasi. Atribut ini diperoleh berdasarkan perhitungan *z-score* dari data usia, jenis kelamin, dan tinggi badan sesuai standar WHO. Nilai *z-score* tersebut kemudian dikategorikan ke dalam tiga kelas, yaitu normal, *stunting*, dan *severely stunting*. Hasil penambahan kolom status dapat dilihat pada tabel 3.2.

Tabel 3.2 Penambahkan kolom Status

No	Nama	Usia bulan	Jenis Kelamin	Tinggi Badan	Status
1	Mizan	24	Laki - Laki	73	*Severely Stunting*
2	Rafa	21	Laki - Laki	78	Normal
3	Raizi	18	Laki - Laki	79.5	Normal
4	Elvano	20	Laki - Laki	88.8	Normal
5	David	19	Laki - Laki	81.3	Normal
...	…	…	…	…	…
505	Ciyra	34	Perempuan	79.7	*Severely Stunting*

3.5.2 *Label encoding*

Label Encoding digunakan untuk mengonversi *variabel* kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Kolom yang di konversi yaitu Jenis kelamin dan Status. Pelabelan kolom status di dapatkan dari perhitungan *z-score* dari standar WHO.

Langkah – Langkah *label encoding* untuk Kolom Jenis Kelamin *:*

1. Mulai
2. Persiapkan Data: kolom "Jenis Kelamin" yang berisi kategori seperti "Laki-laki" dan "Perempuan".
3. Tentukan angka untuk menggantikan kategori:
   "Laki-laki" diganti dengan 0.
   "Perempuan" diganti dengan 1.
4. Periksa setiap baris data satu per satu:
   Jika data pada baris tersebut adalah "Laki-laki", ubah menjadi angka 0.
   Jika data pada baris tersebut adalah "Perempuan", ubah menjadi angka 1.
5. Setelah Semua Data Diproses, Kolom "Jenis Kelamin" Sekarang Berisi Angka.
6. Selesai.

Tabel 3.3 *Label Encoding* Jenis Kelamin

No	Nama	Usia bulan	Jenis Kelamin	Tinggi Badan	Status
1	Mizan	24	0	73	*Severely Stunting*
2	Rafa	21	0	78	Normal
3	Raizi	18	0	79.5	Normal
4	Elvano	20	0	88.8	Normal
5	David	19	0	81.3	Normal
...	…	…	…	…	…
505	Ciyra	34	1	79.7	*Severely Stunting*

Langkah – Langkah *label encoding* untuk Kolom Status *:*

1. Mulai
2. Persiapkan Data
3. Tentukan angka untuk menggantikan kategori:
   "Normal" diganti dengan → 0
   "*Stunting*" diganti dengan 1
   "*Severely Stunting*" diganti dengan 2
4. Periksa setiap baris data satu per satu:
   Jika data pada baris adalah "Normal", ubah menjadi angka 0.
   Jika data pada baris adalah "*Stunting*", ubah menjadi angka 1.
   Jika data pada baris adalah "*Severely Stunting*", ubah menjadi angka 2.

5. Setelah Semua Data Diproses, Kolom "Status" Sekarang Berisi Angka.
6. Selesai

Tabel 3.4 *Label Encoding* Status

No	Nama	Usia bulan	Jenis Kelamin	Tinggi Badan	Status
1	Mizan	24	0	73	2
2	Rafa	21	0	78	0
3	Raizi	18	0	79.5	0
4	Elvano	20	0	88.8	0
5	David	19	0	81.3	0
...	…	…	…	…	…
505	Ciyra	34	1	79.7	2

3.5.3 *Normalisasi* Data

Normalisasi data adalah teknik *transformasi* yang digunakan untuk mengubah nilai atribut numerik ke dalam rentang tertentu, antara 0 hingga 1. Proses ini bertujuan menyamakan skala data agar algoritma pembelajaran mesin dapat beroperasi dengan lebih optimal. Dalam penelitian ini, metode *Min-Max Scaling* digunakan untuk menyesuaikan nilai data agar berada dalam kisaran yang seragam. Dengan pendekatan ini, perbedaan unit atau rentang nilai antar fitur dalam dataset dapat diminimalkan, sehingga analisis menjadi lebih akurat.

Langkah – Langkah *Normalisasi* :

1. Mulai
2. Persiapkan Data
3. Inisialisasi *MinMaxScaler*.
4. Tentukan nilai minimum (X_*min*) dan maksimum (X_*max*) dari data.
5. Untuk setiap nilai X dalam data:

   Hitung nilai *normalisasi* dengan rumus:

   X' = (X - X_*min*) / (X_*max* - X_*min*)

   Gantikan nilai asli dengan nilai X'.
6. Dataset kini dalam bentuk *normalisasi*.
7. Selesai

Tabel 3.5 Hasil *Normalisasi* Data

No	Nama	Usia bulan	Jenis Kelamin	Tinggi Badan	Status
1	Mizan	0.25	0	0.24	2
2	Rafa	0.18	0	0.33	0
3	Raizi	0.12	0	0.35	0
4	Elvano	0.16	0	0.51	0
5	David	0.14	1	0.38	0
...	…	…	…	…	…
505	Ciyra	0.45	1	0.36	2

3.6 Pemilihan Fitur

Pemilihan fitur bertujuan menyederhanakan data dan meningkatkan kinerja model. Korelasi antar fitur serta dengan *variabel* target dianalisis menggunakan *heatmap*. Melalui proses ini, fitur-fitur yang kurang berkontribusi atau memiliki hubungan yang terlalu kuat dengan fitur lain diputuskan untuk tidak digunakan, guna menghindari *overfitting* dan mendukung kinerja model yang lebih optimal.

3.7 *Balancing* data

Balancing data dilakukan untuk mengatasi masalah jumlah data yang tidak seimbang antar kelas. Jika data tidak seimbang, model cenderung lebih sering memprediksi kelas yang jumlahnya lebih banyak, sehingga performa untuk kelas yang sedikit menjadi buruk. Dalam penelitian ini, balancing dilakukan dengan teknik SMOTE (*Synthetic Minority Over-sampling Technique)*. Teknik ini membuat data baru pada kelas minoritas dengan cara mengambil sampel data lama lalu menambah variasi data baru yang mirip. Dengan begitu, jumlah data di setiap kelas menjadi seimbang.

3.8 Pembagian data

Setelah data diproses melalui tahapan *preprocessing* yang mencakup data *cleaning* dan *transformasi*, langkah selanjutnya memisahkan data menjadi dua bagian utama: *training* set dan *test* set. Pembagian data ini bertujuan untuk melatih model menggunakan sebagian data, sementara sebagian lainnya digunakan untuk menguji performa model.

Langkah – Langkah pembagian data :

1. Mulai
2. Siapkan data yang akan dibagi:
   Dataset X untuk fitur (data *input*).
   Dataset Y untuk label (data *output*).
3. Tentukan *proporsi* pembagian data:
   80% untuk *training* (data latih), 20% untuk *testing* (data uji).
4. Acak urutan data untuk memastikan pembagian acak:
   Acak urutan baris dalam dataset
5. Bagi data menjadi dua bagian:
   Data *training*: Ambil 80% data pertama.
   Data *testing*: Ambil 20% data sisanya.
6. *Output* hasil pembagian:
   Data *training*: (X_*train*, Y_*train*).
   Data *testing*: (X_*test*, Y_*test*).
7. Selesai

3.9 *Implementasi* model KNN dan *Naïve Bayes*

Implementasi model pada penelitian ini menggunakan algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* untuk mengklasifikasikan *stunting* pada balita di Desa Pasirjengkol.

3.9.1 *K-Nearest Neighbor*

KNN algoritma yang bekerja dengan cara mencari K data terdekat dengan data yang ingin diprediksi. Berdasarkan kelas *mayoritas* dari data tetangga terdekat tersebut, prediksi akan dibuat.

Langkah – Langkah algoritma KNN :

1. Mulai
2. Siapkan Data
   a. Data *training* : digunakan untuk melatih model
      X_*train* : fitur (usia, jenis kelamin, dan tinggi badan)
      Y_*train* : Label (Status *stunting* : normal, *stunting*, *severely stunting*)
   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:*

      Jarak = √((nilai_uji1 - nilai_latih1)^2 + (nilai_uji2 - nilai_latih2)^2 + ...)

   b. Urutkan semua data berdasarkan jarak terdekat.

   c. 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 kesmunculan masing-masing kelas (Normal, *Stunting*, *Severely stunting).*

   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

3.9.2 *Naïve Bayes*

Naïve Bayes metode klasifikasi berbasis statistik yang digunakan untuk menghitung kemungkinan suatu data masuk ke dalam kelas tertentu.

Langkah – Langkah algoritma *Naïve Bayes* :

1. Mulai
2. Siapkan data :

   Data *Training* (X_*train* untuk fitur, Y_*train* untuk label).

   Data *Testing* (X_*test* untuk fitur, tanpa label).
3. Hitung *probabilitas* awal untuk setiap kelas

Jumlahkan total data dalam setiap kelas.

Hitung *probabilitas* setiap kelas

P (kelas) = (jumlah data dalam kelas)/(total data)

4. Hitung *probabilitas* setiap fitur berdasarkan kelas.

   Hitung jumlah kemunculan nilai fitur dalam setiap kelas.

   Hitung *probabilitas* setiap fitur muncul dalam kelas tersebut.

5. Prediksi data *testing* :

   Untuk setiap data uji:

   a. Hitung *probabilitas* data tersebut masuk ke masing-masing kelas.
   b. Pilih kelas dengan peluang terbesar sebagai hasil prediksi.

6. Simpan hasil prediksi
7. selesai

3.10 Evaluasi Model

Evaluasi model dalam penelitian ini dilakukan dengan menggunakan *Confusion Matrix* untuk menilai performa model *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* dalam mengklasifikasikan status *stunting* pada balita di Desa Pasirjengkol. *Confusion Matrix* memberikan penjelasan yang lebih rinci mengenai cara model mengelompokkan data ke dalam masing-masing kategori, yaitu normal, *stunting*, dan *severely stunting*.

3.10.1 *Confusion Matrix*

Confusion Matrix metode evaluasi yang digunakan menilai performa model klasifikasi dengan membandingkan hasil prediksi dengan nilai sebenarnya. Dalam konteks penelitian ini status *stunting* pada balita dikategorikan menjadi tiga kelas: normal, *stunting* dan *severely stunting*. *Confusion Matrix* akan menggambarkan bagaimana model KNN dan *Naïve Bayes* mengklasifikasikan data ke dalam kelas-kelas tersebut.