

## BAB III

### METODE PENELITIAN

#### 3.1 Objek Penelitian

Penelitian ini dilakukan di Puskesmas Anggadita, yang berlokasi di Jl. Raya Klari No. 41352, Anggadita, Kecamatan Klari, Karawang, Jawa Barat 41371, Indonesia. Puskesmas ini menyediakan pelayanan kesehatan masyarakat, dengan fokus utama pada ibu dan anak, serta menyimpan data status gizi balita yang mencakup berat badan, tinggi badan, usia, dan status gizi berdasarkan prediksi standar. Lokasi penelitian ini dipilih karena relevansi data yang dimiliki terkait masalah stunting serta peran penting puskesmas dalam upaya pencegahan stunting di tingkat lokal. Sebagai lembaga yang secara langsung menangani dan memantau pertumbuhan balita, Puskesmas Anggadita memberikan gambaran nyata mengenai kondisi lapangan yang menjadi fokus penelitian. Penelitian ini dilaksanakan selama empat bulan sesuai dengan tahapan yang dijelaskan pada Tabel 3.1.

Tabel 3.1 Waktu Penelitian

| No | Kegiatan                             | Bulan 1 |   |   |   | Bulan 2 |   |   |   | Bulan 3 |   |   |   | Bulan 4 |   |   |   |
|----|--------------------------------------|---------|---|---|---|---------|---|---|---|---------|---|---|---|---------|---|---|---|
|    |                                      | 1       | 2 | 3 | 4 | 1       | 2 | 3 | 4 | 1       | 2 | 3 | 4 | 1       | 2 | 3 | 4 |
| 1  | Menentukan Tema                      |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |
| 2  | Studi Literatur                      |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |
| 3  | Pengumpulan Data                     |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |
| 4  | Eksplorasi Data                      |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |
| 5  | Pra-pemrosesan Data                  |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |
| 6  | Pembuatan Model                      |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |
| 7  | Evaluasi Model                       |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |
| 8  | Penulisan Laporan dan Revisi Laporan |         |   |   |   |         |   |   |   |         |   |   |   |         |   |   |   |

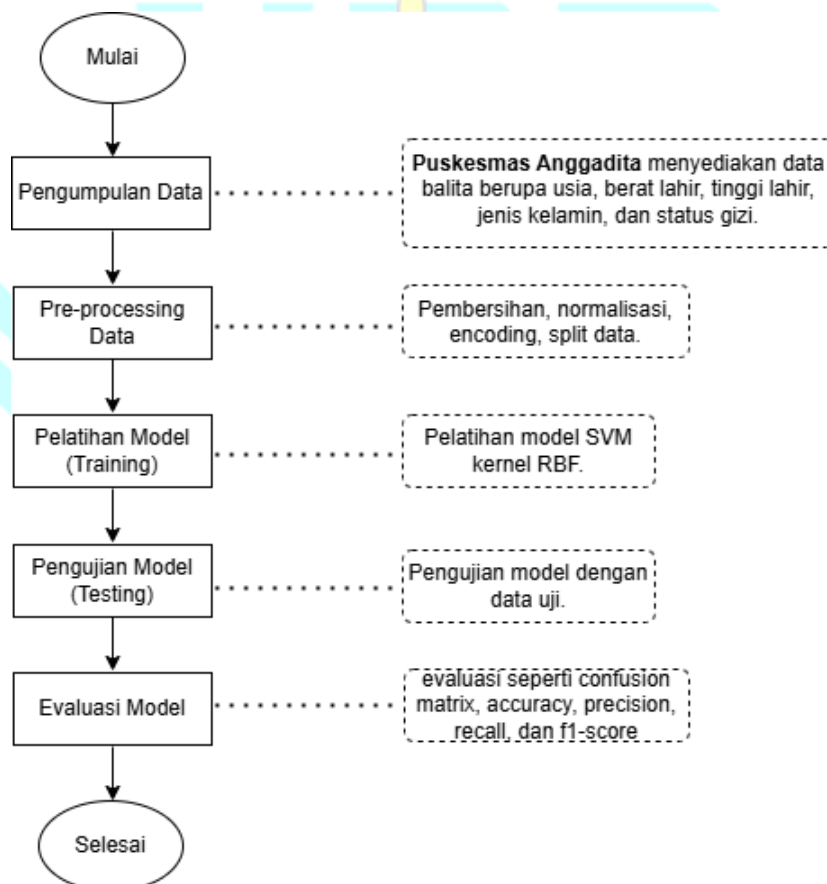
Penelitian ini difokuskan pada data balita yang terdaftar di Puskesmas Anggadita, terutama yang berhubungan dengan status gizi masyarakat. Penelitian ini menggunakan metode ilmu data berbasis data untuk menghasilkan analisis yang akurat dan dapat diandalkan. Selain itu, penelitian ini mengkaji hubungan antara variabel-variabel seperti usia, berat badan, tinggi badan, status gizi, dan faktor lainnya untuk melihat korelasi yang signifikan. Analisis hanya mencakup variabel

yang relevan dalam memprediksi risiko stunting serta didasarkan pada data yang tersedia di lokasi penelitian.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengambilan kebijakan kesehatan masyarakat, khususnya yang berkaitan dengan pencegahan dan penanganan risiko angka stunting di Kabupaten Karawang dan wilayah sekitarnya. Dengan demikian, penelitian ini tidak hanya memberikan wawasan mengenai faktor-faktor yang memengaruhi stunting.

### 3.2 Prosedur Penelitian

Prosedur penelitian ini dimulai dengan studi literatur dan diakhiri dengan evaluasi performa model. Berikut adalah flowchart yang menggambarkan alur lengkap dari tahap-tahap penelitian, mulai dari pengumpulan data, pelatihan model, hingga evaluasi hasil prediksi stunting:



Gambar 3.1 Flowchart Alur Penelitian

### 3.3 Pengumpulan Data

Pengumpulan data dalam penelitian ini bertujuan untuk memprediksi risiko stunting pada balita menggunakan algoritma *Support Vector Machine*. Data diperoleh dari Puskesmas Anggadita Karawang, dengan cakupan informasi seperti identitas balita, usia, jenis kelamin, berat badan lahir, tinggi badan lahir, serta indikator BB/U (Berat Badan per Umur) dan TB/U (Tinggi Badan per Umur). Selain itu, data juga mencakup informasi mengenai pola makan balita, status gizi ibu, dan riwayat kesehatan keluarga. Data yang digunakan merupakan data stunting pada bulan Januari hingga Desember, yang berjumlah sebanyak 1.028 data balita dari satu kecamatan.

Setelah data terkumpul, langkah selanjutnya adalah menyusunnya dalam format terstruktur, seperti CSV atau Excel, untuk mempermudah proses pengolahan dan analisis lebih lanjut. Fitur-fitur yang dikumpulkan kemudian disusun dalam bentuk tabel yang terorganisir untuk mendukung proses prediksi risiko stunting secara akurat dan efisien. Berikut data stunting yang digunakan:

Tabel 3.2. Data Stunting

|    | Nama                          | Usia Saat Ukur              | BB Lahir | TB Lahir | JENIS KELAMIN | BB/U               | TB/U          | Status Gizi |
|----|-------------------------------|-----------------------------|----------|----------|---------------|--------------------|---------------|-------------|
| 2  | AISYAH                        | 3 Tahun - 5 Bulan - 26 hari | 03.04    | 50       | Perempuan     | Kurang             | Normal        | Stunting    |
| 3  | BUDI                          | 4 Tahun - 8 Bulan - 27 Hari | 02.04    | 49       | Laki-laki     | Normal             | Normal        | Normal      |
| 4  | BARKA FAUZI                   | 4 Tahun - 9 Bulan - 29 Hari | 3        | 50       | Laki-laki     | Berat Badan Normal | Pendek        | Normal      |
| 5  | JIBRIL MAULANA                | Hari                        | 3        | 50       | Laki-laki     | Kurang             | Pendek        | Normal      |
| 6  | MICHELLE P A                  | 4 Tahun - 7 Bulan - 27 Hari | 03.02    | 51       | Laki-laki     | Kurang             | Pendek        | Normal      |
| 7  | GIO HENDRA RIZKY<br>RAMADHAN  | 4 Tahun - 3 Bulan - 12 Hari | 03.01    | 51       | Laki-laki     | Berat Badan Normal | Pendek        | Normal      |
| 8  | AISHAWA MIKHAYLA<br>NAFISAH   | 4 Tahun - 9 Bulan - 1 Hari  | 03.02    | 52       | Perempuan     | Berat Badan Normal | Pendek        | Normal      |
| 9  | HAWA PUTRI BAHTIAR            | 3 Tahun - 10 Bulan - 0 Hari | 3        | 50       | Perempuan     | Berat Badan Normal | Pendek        | Normal      |
| 10 | AHMAD BANG BANG<br>AMAD IKSAN | 3 Tahun - 8 Bulan - 29 Hari | 02.06    | 48       | Laki-laki     | Kurang             | Sangat Pendek | Normal      |
| 11 | NAJYAN                        | 4 Tahun - 5 Bulan - 28 Hari | 03.02    | 48       | Laki-laki     | Kurang             | Pendek        | Normal      |
| 12 | RENO NOVIANSYAH               | 4 Tahun - 8 Bulan - 27 Hari | 02.08    | 48       | Laki-laki     | Kurang             | Pendek        | Normal      |

### 3.4 Pre- pemrosesan Data

Pada tahap pra-pemrosesan data, beberapa langkah penting dilakukan untuk memastikan dataset dalam kondisi bersih dan layak digunakan dalam proses prediksi risiko stunting. Pertama, dilakukan pemeriksaan nilai kosong (missing values) menggunakan perintah `data.isnull().sum()` untuk mengetahui jumlah data kosong di setiap kolom. Nilai kosong pada variabel numerik seperti berat

badan, tinggi badan, dan usia diisi menggunakan metode imputasi rata-rata (mean imputation) dengan perintah `fillna(data['Nama Kolom'].mean(), inplace=True)`, sedangkan untuk variabel kategorikal seperti jenis kelamin dan status gizi diisi menggunakan modus (mode imputation) melalui perintah `fillna(data['Nama Kolom'].mode()[0], inplace=True)`. Setelah itu, data duplikat dideteksi menggunakan perintah `data.duplicated().sum()` dan dihapus dengan `data.drop_duplicates(inplace=True)` agar setiap data yang digunakan bersifat unik dan tidak terjadi pengulangan yang dapat memengaruhi hasil analisis.

Selanjutnya, untuk mendeteksi dan menangani anomali (outlier) digunakan metode Interquartile Range (IQR) dengan rumus  $IQR = Q3 - Q1$ , Dimana data yang memiliki nilai di bawah  $Q1 - 1.5 \times IQR$  atau di atas  $Q3 + 1.5 \times IQR$  dianggap sebagai outlier. Proses ini dilakukan dengan perintah `quantile()` untuk menghitung Q1 dan Q3, lalu membuat filter outlier dan menghapus data yang masuk kategori anomali. Setelah tahap pembersihan selesai, variabel numerik dinormalisasi menggunakan teknik Min-Max Scaling atau Standard Scaling agar skala antar fitur konsisten. Variabel kategorikal seperti status gizi juga diubah menjadi format numerik menggunakan Label Encoding, di mana kategori "Stunting" diberi nilai 1 dan "Tidak Stunting" diberi nilai 0. Terakhir, data dibagi menjadi training set dan testing set dengan rasio 80:20 menggunakan teknik stratified sampling untuk menjaga proporsi seimbang antar kelas dalam kedua dataset tersebut.

### 3.5 Pembuatan Model

Tahap pembuatan model dilakukan setelah data selesai diproses dalam tahap pra-pemrosesan. Model *Support Vector Machine* dibangun menggunakan dataset pelatihan yang telah melewati tahap pembersihan dan transformasi data. Pada penelitian ini, model SVM menggunakan kernel Radial Basis Function (RBF), karena kernel ini memiliki kemampuan yang baik dalam menangani data dengan distribusi non-linear dan kompleks, seperti data balita yang memiliki variabilitas tinggi pada faktor-faktor berat badan, tinggi badan, usia, serta status gizi. Pemilihan kernel menjadi langkah penting dalam tahap ini, karena kernel menentukan cara SVM memisahkan data dalam ruang dimensi yang lebih

tinggi. Beberapa jenis kernel yang umum digunakan antara lain linear, polynomial, dan radial basis function (RBF), namun berdasarkan uji awal dan karakteristik pola data yang dianalisis, kernel RBF dipilih karena lebih adaptif terhadap pola distribusi data yang tidak linear. Tabel 3.3 Pseudocode Prediksi Risiko Stunting dengan SVM.

Selain pemilihan kernel, tahap ini juga mencakup optimasi parameter model, seperti nilai regulasi (C) dan parameter gamma, yang dapat memengaruhi kompleksitas serta performa model. Optimasi parameter dilakukan menggunakan teknik seperti Grid Search atau Random Search untuk menemukan kombinasi terbaik yang menghasilkan model dengan performa tinggi. Selama proses pelatihan, dilakukan validasi silang (cross-validation) guna mengukur kemampuan model dalam menggeneralisasi data dan menghindari masalah overfitting. Setelah model awal terbentuk, dilakukan evaluasi terhadap performa model dengan melihat bagaimana model memahami pola data dari dataset pelatihan. Jika hasil yang diperoleh belum optimal, maka dilakukan perbaikan dengan penyesuaian parameter atau penambahan data pelatihan. Model yang telah memenuhi kriteria evaluasi kemudian akan diuji lebih lanjut menggunakan dataset pengujian sebelum akhirnya digunakan dalam prediksi risiko stunting pada balita.

Tabel 3.3 Pseudocode Prediksi Risiko Stunting dengan SVM

|                                                                                  |
|----------------------------------------------------------------------------------|
| <p>Algoritma : Prediksi Risiko Stunting dengan <i>Support Vector Machine</i></p> |
|----------------------------------------------------------------------------------|

Input :

- Data Latih (X\_Train, Y\_Train) : Data latih digunakan untuk membangun model. X\_Train berisi fitur-fitur seperti usia, berat badan, dan pola makan, sedangkan Y\_Train adalah label yang menunjukkan status stunting.
- Data Uji (X\_Test) : Data uji digunakan untuk mengukur performa model setelah pelatihan. Model akan memprediksi label berdasarkan fitur dalam X\_Test, lalu hasilnya dibandingkan dengan data sebenarnya.
- Pemilihan Kernel (Linear, RBF, Polynomial, atau Sigmoid) : Kernel menentukan cara model memisahkan data. Linear cocok untuk data

---

terpisah garis lurus, RBF untuk pola kompleks, Polynomial untuk hubungan non-linear, dan Sigmoid lebih jarang digunakan dalam SVM.

- *Hyperparameter* (C, Gamma, Degree) : C mengontrol keseimbangan antara akurasi dan generalisasi, Gamma menentukan pengaruh satu titik data terhadap lainnya (pada RBF/Polynomial), dan Degree hanya untuk Polynomial Kernel, menentukan kompleksitas pemisahan data.

Output :

1. Membaca Data Latih : Dataset latih dimuat untuk digunakan dalam proses pelatihan model. Data ini berisi fitur-fitur yang akan digunakan untuk membangun hubungan antara variabel input dan target.
2. Membaca Data Uji : Dataset uji dimuat untuk mengukur performa model setelah pelatihan. Data ini tidak digunakan saat pelatihan agar evaluasi lebih objektif.
3. Melakukan Pra-Pemrosesan Data : Proses ini mencakup pembersihan, transformasi, dan penyesuaian data agar siap digunakan dalam model. Langkah ini penting untuk memastikan akurasi model.
4. Menangani nilai yang hilang dengan imputasi (mean/median/modus) : Jika terdapat data yang hilang, metode imputasi diterapkan. Mean digunakan untuk data numerik dengan distribusi normal, median untuk data dengan outlier, dan modus untuk variabel kategorikal.
5. Menghapus atau menangani outlier menggunakan metode *IQR* atau *Z-score* : Outlier dideteksi menggunakan metode Interquartile Range (IQR) atau *Z-score*. Data ekstrem dapat dihapus atau diubah agar tidak mempengaruhi performa model.
6. Encoding variabel kategorikal (misalnya, jenis kelamin) : Variabel kategorikal seperti jenis kelamin dikonversi menjadi bentuk numerik agar dapat dipahami oleh model, menggunakan metode seperti Label Encoding atau *One-Hot Encoding*.
7. Normalisasi/standarisasi fitur numerik agar memiliki skala yang seimbang : Data numerik disesuaikan ke skala yang sama menggunakan normalisasi (Min-Max Scaling) atau standarisasi (Z-score

---

normalization), agar model tidak berat sebelah pada fitur dengan skala besar.

8. Membagi Dataset menjadi Data Latih dan Data Uji (misalnya 80% latih, 20% uji) : Dataset dibagi menjadi dua bagian: data latih (80%) untuk melatih model dan data uji (20%) untuk mengukur performanya. Pembagian ini memastikan model tidak hanya menghafal data tetapi juga dapat melakukan generalisasi dengan baik.

### 3.6 Evaluasi Model

Evaluasi model merupakan tahap penting untuk menilai kinerja algoritma *Support Vector Machine* (SVM) dalam memprediksi risiko stunting pada balita. Evaluasi ini dilakukan dengan menggunakan data uji yang telah dipisahkan sebelumnya dari dataset utama. Tujuan utama dari tahap ini adalah mengukur sejauh mana model mampu menggeneralisasi pola data baru yang belum pernah dilihat sebelumnya.

Metode evaluasi yang digunakan dalam penelitian ini melibatkan beberapa metrik utama, yaitu akurasi (*accuracy*), presisi (*precision*), recall (sensitivitas), F1-score, serta confusion matrix. Metrik ini digunakan untuk menilai performa model secara menyeluruh. Confusion matrix memberikan gambaran lebih detail mengenai jumlah prediksi yang benar dan salah dalam masing-masing kategori, seperti True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari confusion matrix ini, nilai precision, recall, dan F1-score dapat dihitung untuk memahami keseimbangan antara ketepatan dan sensitivitas model.

Selain itu, dilakukan validasi silang (*cross-validation*) untuk mengukur kestabilan model dan menghindari kemungkinan *overfitting*. Teknik yang digunakan dalam validasi silang adalah *K-Fold Cross Validation*, yang membagi dataset menjadi beberapa bagian untuk memastikan bahwa model diuji pada berbagai subset data. Dengan cara ini, model diuji secara lebih menyeluruh dan hasil evaluasi menjadi lebih akurat serta tidak hanya bergantung pada satu skenario pembagian dataset.

Jika hasil evaluasi menunjukkan performa yang kurang optimal, dilakukan iterasi ulang dengan penyesuaian *hyperparameter* tuning. Parameter seperti nilai C, gamma, dan jenis kernel (Linear, RBF, Polynomial, atau Sigmoid) dapat dioptimalkan menggunakan metode seperti *Grid Search* atau Random Search. Selain itu, dapat dilakukan penyesuaian pada fitur yang digunakan dalam model dengan menerapkan teknik seleksi fitur untuk menghilangkan fitur yang kurang berkontribusi dalam prediksi.

Hasil evaluasi akhir akan digunakan untuk menentukan apakah model SVM yang dikembangkan sudah layak untuk diterapkan dalam skenario nyata atau masih memerlukan perbaikan. Jika model mencapai tingkat akurasi dan F1-score yang tinggi, maka dapat disimpulkan bahwa model memiliki kemampuan prediksi yang baik dalam mengidentifikasi balita yang berisiko mengalami stunting.

