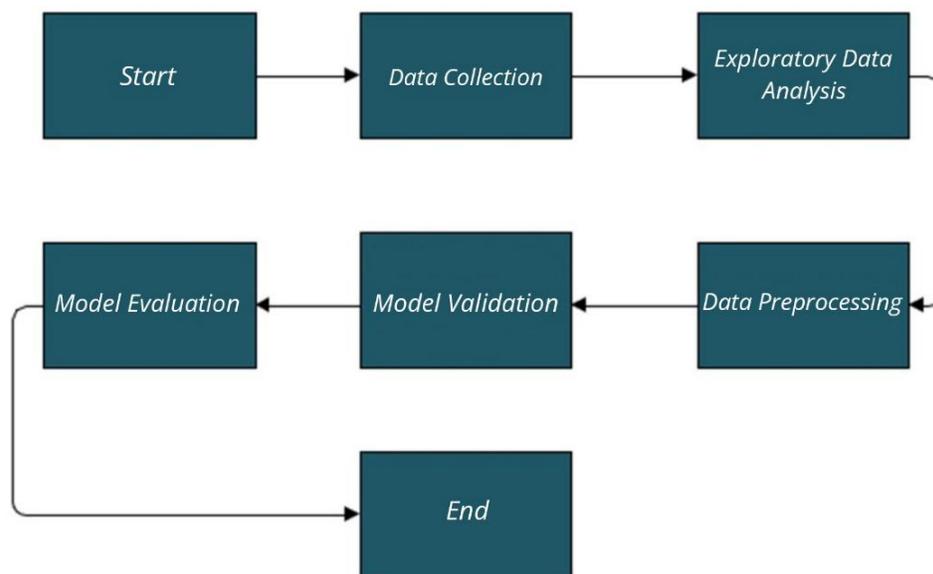


## BAB III METODOLOGI PENELITIAN

### 3.1 Alur Penelitian

Metode penelitian yang digunakan pada penelitian ini menggunakan 3 algoritma yaitu *Decision Tree*, *Naïve Bayes* dan *Support Vector Machine*. Adapun tahapan penelitian yang dilakukan oleh penulis sebagai berikut :



**Gambar 3.1** Alur Penelitian

#### 3.1.1 Data Collection

Penelitian ini menggunakan dataset resiko kasus stunting pada balita yang diperoleh dari *Kaggle repository*. Dataset yang digunakan berjumlah 7.573 data dalam bentuk comma separated values (csv). Pada data tersebut terdiri dari 8 column untuk Kolom sebagai predicted adalah gender, age, birth weight, birth length, body weight, body length, dan breastfeeding, Sedangkan column

sebagai target adalah column stunting. Berikut contoh 5 record dari dataset penelitian :

**Tabel 3.1** Data Sample

	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Breastfeeding	Stunting
0	female	56	2.9	50	11.0	90.0	Yes	No
1	female	20	3.3	49	11.1	80.5	No	No
2	male	4	2.8	48	6.5	63.0	No	No
3	female	14	2.0	49	7.0	71.0	Yes	No
4	male	32	3.2	49	11.0	88.7	Yes	No

**Tabel 3.2** Metadata

<i>Attribute</i>	<i>Description</i>	<i>Value</i>	<i>Type Data</i>
gender	Jenis Kelamin anak	<i>Male dan Female</i>	<i>Categorical</i>
age	Usia anak (dalam bulan))	0 - 48	<i>Numerical</i>
Birth Weight	Berat lahir (KG)	2.0 – 3.1	<i>Numerical</i>
Birth Length	Panjang lahir (cm)	48 - 50	<i>Numerical</i>
Body Weight	Berat badan anak saat ini (KG)	2.9 – 10.5	<i>Numerical</i>
Body Length	Panjang badan anak saat ini	49.0 – 92.7	<i>Numerical</i>

Breastfeeding	Status Menyusui	Yes atau No	<i>Categorical</i>
Stunting	Status stunting ( 0 = No, 1 = Yes)	0 atau 1	<i>Categorical</i>

### 3.1.2 *Exploratory Data Analysis (EDA)*

Dalam penelitian ini terdiri dalam beberapa langkah pengerjaan yaitu tahapan analyzing dataset kemudian dilanjutkan dengan deskripsi dataset menggunakan EDA (Exploratory Data Analysis), dengan melakukan EDA, akan sangat berguna dalam mendeteksi kesalahan dari awal, dapat mengidentifikasi outlier, mengetahui hubungan antar data serta dapat menggali faktor-faktor penting dari data [22]. Data yang telah dikumpulkan akan dianalisis terlebih dahulu untuk mengetahui karakteristik data yang memiliki nilai yang duplikat, null, jumlah data dan tipe data yang tidak konsisten, fungsi lain EDA adalah mengenali kesalahan dataset dengan menguasai pola suatu data dan menemukan hubungan antar variabel.[22]

Dalam melakukan analisa data yang terkait dengan resiko stunting pada balita dengan menerapkan Exploratory data analysis (EDA) seperti teknik visualisasi sebagai berikut :

a. Memahami Data

Proses EDA dimulai dengan memahami struktur dan tipe data yang dimiliki. yaitu mencakup pemeriksaan bentuk dataset, jenis variabel (numerik atau kategori) serta melihat ringkasan statistik dari dataset. Pemahaman ini penting untuk memberikan gambaran awal tentang karakteristik dataset yang akan dianalisis.

b. *Duplicate*

Dilakukan pemeriksaan pada dataset untuk menghindari bias pada model yang dapat mempengaruhi representasi data yang sebenarnya. Data duplikat juga dapat menyebabkan bias pada model dan memengaruhi akurasi prediksi, sehingga penting untuk memastikan bahwa setiap baris data bersifat unik. Setelah dilakukan pemeriksaan, diketahui bahwa dataset tidak mengandung data duplikat dan jumlah record tetap 7573 data.

c. Identifikasi *Missing Values*

memeriksa apakah terdapat nilai yang hilang atau null dan nilai duplikat pada data, untuk itu diperlukan identifikasi missing values untuk meningkatkan akurasi kinerja model guna meningkatkan keakuratan data, setelah dilakukan pengecekan dataset tidak memiliki *missing value* membuat jumlah dataset tetap dengan jumlah awal yaitu 7573 record data.

d. *Outlier*

Pada penelitian ini pemeriksaan outlier menggunakan visualisasi boxplot dan IQR ( Inter Quartile Range) digunakan untuk mendeteksi outlier dalam data dengan cara menentukan batas bawah (lower) dan batas atas (upper) di luar mana data dianggap sebagai outlier. Setelah dilakukan pengecekan outlier jumlah record dalam dataset berubah menjadi 4480 record data.

e. *Univariate Analysis*

Pada tahap analisis univariate, saya melakukan beberapa langkah untuk memahami distribusi dan karakteristik dari masing-masing variabel dalam dataset. Langkah pertama adalah melakukan eksplorasi data secara deskriptif untuk mendapatkan gambaran mengenai jumlah sampel, nilai minimum, maksimum, rata-rata, serta proporsi pada variabel kategorik maupun numerik. Selanjutnya, saya membuat visualisasi berupa grafik untuk mempermudah interpretasi hasil analisis.

f. *Multivariate Analysis*

Pada tahap analisis multivariat, saya menganalisis hubungan antarvariabel numerik dalam dataset, yaitu Age, Birth Weight, Birth Length, Body Weight, dan Body Length, menggunakan pairplot. Pairplot menampilkan scatter plot untuk hubungan antarvariabel dan distribusi univariat di diagonal.

### **3.1.3 Data Preprocessing**

Pada tahapan ini akan dilakukan persiapan data yaitu :

a. Normalisasi Data

Normalisasi data dilakukan untuk memastikan bahwa semua fitur memiliki skala yang sama, pada penelitian ini penulis menggunakan z-score untuk proses normalisasi data.

b. Encoding fitur kategori

Proses mengubah suatu fitur yang memiliki nilai kategori, dimana nilai dari setiap kolom akan diganti menjadi angka. Pada tahapan ini adanya perubahan format numerik agar dapat digunakan dalam algoritma mesin. Dengan mengubah kategori menjadi 0 dan 1 pada kolom target yaitu stusting.[7]

c. Data Splitting

proses memisahkan data latih dan data uji, yaitu dataset menjadi dua bagian yaitu data latih (training) dan data uji (testing). Data latih digunakan untuk melatih model agar mengenali pola dan hubungan dalam data, Dalam penelitian ini proses split ini ditentukan data training 80% dan data testing 20%. Langkah ini penting agar model dapat menggeneralisasi dengan baik dan tidak hanya berfokus pada data latih, sehingga menghasilkan prediksi yang akurat saat digunakan. Pada penelitian ini jumlah data training sebanyak 3584 dan data testing sebanyak 896[15].

### 3.1.4 Model Validation

Validasi model dilakukan untuk mengevaluasi performa algoritma yang digunakan dalam penelitian ini, yaitu *Decision Tree*, *Naive Bayes*, dan *Support Vector Machine*. Langkah-langkah validasi yang dilakukan adalah sebagai berikut:

1. Training

Proses pelatihan model melibatkan penyesuaian parameter untuk meminimalkan error dengan menggunakan data pelatihan untuk menemukan pola atau hubungan antara fitur dan target berjumlah 3584 record data. Setelah model dilatih, metrik evaluasi digunakan untuk menilai performa model pada data pelatihan dan data pengujian.

2. Cross Validation

Pada tahap Model Validation, dilakukan evaluasi terhadap kinerja model menggunakan teknik 5-fold cross-validation dengan proporsi pembagian data 80-20. Pertama, dataset dibagi menjadi lima bagian yang hampir sama besar. Sebanyak 80% data digunakan untuk pelatihan model, sedangkan 20% sisanya digunakan untuk pengujian pada setiap fold. Proses ini diulang sebanyak lima

kali, dengan setiap bagian data menjadi data uji sekali, sementara empat fold lainnya digunakan sebagai data pelatihan. Dengan cara ini, model diuji secara menyeluruh pada berbagai bagian dataset, sehingga memberikan gambaran yang lebih stabil dan mengurangi bias yang mungkin timbul akibat pembagian data yang tidak representatif.

### 3. Metrik Evaluasi

Untuk mengukur performa setiap model, digunakan beberapa metrik evaluasi sebagai berikut:

- Akurasi: Mengukur persentase prediksi yang benar terhadap total data testing.
- Precision: Tingkat keakuratan model dalam memprediksi kelas positif (kasus stunting).
- Recall: Kemampuan model dalam mendeteksi seluruh kasus positif (sensitivitas).
- F1-Score: Kombinasi dari precision dan recall untuk menangani ketidakseimbangan data.

#### 3.1.5 *Model Evaluation*

Pada fase ini dilakukan proses pembentukan keluaran yang mudah dimengerti, dimana tahapan ini melakukan penilaian kinerja model dan memastikan model dengan mengukur performa model dengan menggunakan confusion matrix, yang dapat membantu mengidentifikasi model dalam memprediksi kelas.

#### 3.1.6 **Perhitungan Manual**

[https://drive.google.com/drive/folders/1HdrlHxdJdjMA1-o-HmMsscAyD\\_gwhSk?usp=drive\\_link](https://drive.google.com/drive/folders/1HdrlHxdJdjMA1-o-HmMsscAyD_gwhSk?usp=drive_link)