BAB 3
METODE PENELITIAN

Penelitian ini adalah jenis penelitian *kuantitatif*. Menurut KBBI, *kuantitatif* artinya berdasarkan jumlah atau banyaknya. Mengambil data dalam jumlah banyak merupakan pengertian dari penelitian *kuantitatif*. Sedangkan metode penelitian yang dilakukan berupa analisis klasifikasi tentang balita yang terindikasi *stunting* di Puskesmas Kebumen II berdasarkan data dari Puskesmas Kebumen II.

3.1 BAHAN DAN ALAT PENELITIAN

Penelitian ini membutuhkan bahan yang akan digunakan yaitu data *stunting* di Puskesmas Kebumen II selama 3 bulan tahun 2024. Unsur data yang dibutuhkan yaitu jenis kelamin, usia, tinggi badan, berat badan, lingkar kepala, berat badan saat lahir, dan faktor yang menyebabkan terjadinya *stunting*. Data-data tersebut didapatkan dari Puskesmas Kebumen II.

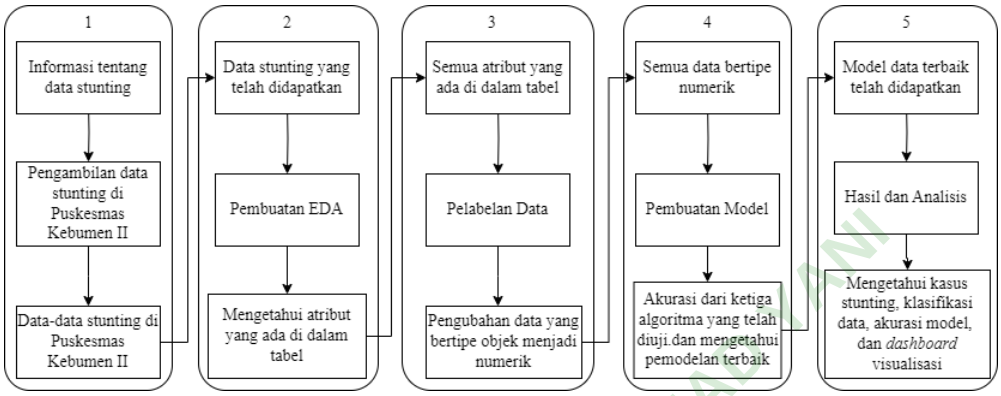
Penelitian yang akan dilakukan membutuhkan beberapa alat yang akan digunakan seperti komputer dengan spesifikasi cukup untuk menjalankan sistem operasi yang dibutuhkan dalam membuat dan merancang perangkat lunak serta koneksi internet yang memadai.

Berikut adalah sistem operasi dan program-program aplikasi yang digunakan dalam pengembangan aplikasi ini:

1. Sistem Operasi Windows 10 atau lebih baru.
2. Bahasa Pemrograman *Python* versi 3.11.3 atau yang terbaru.
3. Microsoft Excel 2013 atau yang lebih baru.
4. *Jupyter Notebook* versi 6.4.12 dan *Sublime Text*.

3.2 JALAN PENELITIAN

Jalan penelitian yang akan dilakukan serta tahapan tahapannya ditunjukkan pada Gambar 3.1 sebagai berikut.



Gambar 3.1 Jalan Penelitian

3.2.1 Tahap Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa *dataset* tentang balita yang diperoleh dari Puskesmas Kebumen II. Data yang didapatkan adalah data balita selama 4 bulan yaitu bulan Januari sampai bulan April pada tahun 2024. Data pada bulan Januari sebanyak 2.759 data, bulan Februari sebanyak 2.746 data, dan bulan Maret 2.567 data. Data balita yang telah diperoleh dapat dilihat pada Tabel 3.1.

Tabel 3.1 Data Balita Puskesmas

No	JK	Usia	Tinggi	Berat	TB_Umur	Posyandu
1.	P	326	77.5	8.6	Normal	MUGI MAKMUR
2.	P	61	55	3.9	Normal	LARASATI 2
3.	L	157	65	6.7	Normal	LARASATI 2
4.	P	609	78	8.5	Normal	LARASATI 2
5.	P	117	60	5.3	Normal	LARASATI 2
6.	L	126	62.5	6.8	Normal	LARASATI 2
7.	L	399	73	7.8	Normal	LARASATI 2
8.	L	566	74.8	7.6	Pendek	LARASATI 2
9.	L	656	80	9.8	Normal	Dewi Sri
10.	P	906	83.5	10.8	Normal	LARASATI 2

3.2.2 Tahap Pembuatan EDA (Exploratory Data Analysis)

EDA adalah salah satu proses yang dilakukan sebelum menganalisis suatu data atau membuat pemodelan. EDA digunakan untuk mengetahui dan memahami atribut yang ada pada *dataset*. Langkah-langkah yang dilakukan dalam pembuatan EDA adalah memeriksa kolom, melihat ukuran data, menampilkan informasi, menampilkan data kosong, menampilkan deskriptif data, mengecek data *duplicate*, dan mengelompokkan data.

Menampilkan kolom bertujuan untuk melihat kolom apa saja yang terdapat pada data yang digunakan. Kolom yang digunakan pada data ini berupa kolom nama, jenis kelamin, usia, tinggi, badan, tinggi badan berdasarkan umur, berat badan berdasarkan usia, berat badan berdasarkan tinggi badan, posyandu, dan alamat. Melihat ukuran data bertujuan untuk memahami ukuran data yang akan digunakan. Data yang akan digunakan sebanyak 2759 data dan panjang baris kolom sebanyak 10 data. Kemudian, menampilkan informasi digunakan untuk mengetahui informasi detail berupa tipe data yang ada pada data yang digunakan. Tipe data tersebut yaitu object, integer, dan float dengan data yang memiliki tipe data float sejumlah 1 data, integer berjumlah 1 data, dan object berjumlah 8 data. Kemudian, menampilkan data kosong yang berfungsi untuk mengetahui kolom mana saja yang memiliki data yang kosong. Langkah selanjutnya adalah mencari deskriptif data. Deskriptif data digunakan untuk mengetahui ringkasan data statistik dan numerik dalam data yang digunakan. Ringkasan tersebut meliputi jumlah, rata-rata, standar deviasi, nilai minimum, kuartil pertama atau nilai di bawah 25%, median (nilai di bawah 50%), kuartil ketiga atau nilai di bawah 75%, dan nilai maksimal dalam setiap kolom.

Data yang digunakan biasanya memiliki kesamaan baik yang identik sama maupun yang hampir sama. Oleh karena itu, data tersebut perlu dicek untuk mengetahui apakah terdapat kolom atau baris yang sama dalam sebuah data. Karena pengecekan data merupakan langkah penting dalam pembuatan EDA dan analisis data untuk selanjutnya digunakan dalam mencari akurasi. Kode untuk mengecek data yang *duplicate* sebagai berikut.

```
#Cek Data Duplicate
df.duplicated().sum()
```

Output yang diperoleh dari langkah tersebut yaitu tidak terdapat data yang memiliki kesamaan (*duplicated*). Selanjutnya adalah pengelompokkan data. Pengelompokkan data digunakan untuk analisis data yang membutuhkan kategori atau suatu kelompok. Pengelompokkan data dapat dilakukan berdasarkan satu atau lebih kolom. Kode untuk mengelompokkan data sebagai berikut.

```
#Mengelompokan data
df.groupby('TB_Umur').mean()
```

Kolom yang dikelompokkan berupa kolom tinggi badan umur. *Output* yang diperoleh dari langkah yang telah dijalankan ditunjukkan pada Tabel 3.2.

Tabel 3.2 *Output* Pengelompokkan Data

TB_Umur	Usia	Berat
Normal	935.417258	11.510120
Pendek	1049.233129	10.008620
Sangat Pendek	993.571429	9.285357
Tinggi	851.000000	15.100000

3.2.3 Tahap Pelabelan Data

Pada tahap ini, pelabelan data dilakukan untuk membuat pembelajaran pada mesin. Pelabelan ini diberi nilai 0 dan 1. Balita yang terindikasi *stunting* akan diberi label 1 atau positif sedangkan balita normal (tidak terindikasi *stunting*) akan diberi label 0 atau negatif. Pelabelan data menjadi suatu proses yang sangat penting karena jika ada yang tidak tepat maka akan mempengaruhi kinerja dari mesin tersebut dan prediksi mesin pada data baru menjadi tidak akurat. Data yang diberi pelabelan terlihat pada Tabel 3.3.

Tabel 3.3 Data yang Sudah Diberi Label

No	JK	Usia	Tinggi	Berat	TB_Umur	Posyandu	Label
1.	P	326	77.5	8.6	Normal	MUGI MAKMUR	0

2.	P	61	55	3.9	Normal	LARASATI 2	0
3.	L	157	65	6.7	Normal	LARASATI 2	0
4.	P	609	78	8.5	Normal	LARASATI 2	0
5.	P	117	60	5.3	Normal	LARASATI 2	0
6.	L	126	62.5	6.8	Normal	LARASATI 2	0
7.	L	399	73	7.8	Normal	LARASATI 2	0
8.	L	566	74.8	7.6	Pendek	LARASATI 2	1
9.	L	656	80	9.8	Normal	Dewi Sri	0
10.	P	906	83.5	10.8	Normal	LARASATI 2	0

Dari jumlah data sebanyak 500 data yang akan digunakan untuk proses *training* dan *testing*, data yang diberi label 0 berjumlah 458 data dan untuk label 1 berjumlah 42 data

3.2.4 Tahap Pembuatan Model

Di tahap pembuatan model, data akan dibagi menjadi 2 yaitu data *training* dan data *testing*. Data *training* digunakan untuk melakukan prediksi secara otomatis sedangkan data *testing* digunakan untuk mengetahui keakuratan model data yang dibuat. Pembuatan model menggunakan beberapa *library* yaitu *numpy, pandas, matplotlib, seaborn, sklearn, dan pickle.* Tahap pembuatan model data ini menggunakan algoritma KNN. Berikut langkah-langkah dalam pembuatan model.

1. Membuat nilai X dan y

Di dalam pembuatan model, perlu dipisahkan antara nilai X dan y. Nilai X disebut juga dengan fitur sementara y disebut dengan variabel target. Nilai X digunakan untuk melatih mesin atau *input* untuk model sedangkan y digunakan untuk *output* atau prediksi suatu target. Kode untuk memisahkan nilai X dan y sebagai berikut.

```
#Membuat nilai X dan y
X = df[['JK', 'Tinggi', 'Usia']]
y = df['Kelas']
```

Untuk membuat nilai X membutuhkan kolom ‘JK, Tinggi, dan Usia’. Kolom-kolom tersebut yang nantinya akan digunakan mesin sebagai acuan pembelajaran.

Sedangkan nilai y diisi dengan kolom ‘Kelas’. Karena kolom tersebut adalah variabel yang ingin diprediksi.

2. *Splitting* Data

Dataset yang akan digunakan dibagi menjadi data *training* dan *testing*. Data *training* digunakan sebagai data latih dalam model sedangkan data *testing* digunakan sebagai data uji untuk menguji seberapa akurat model yang dibuat.

```
#Membagi data training dan data testing
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
```

Kode ‘X_train’ dan ‘y_train’ digunakan untuk melatih model sedangkan ‘X_test’ dan ‘y_test’ digunakan untuk menguji pemodelan yang telah dibuat. Fungsi dari kode ‘train_test_split adalah untuk pembagian *dataset* menjadi data *training* dan *testing*. Fungsi tersebut berasal dari *library* scikit-learn (sklearn). Parameter ‘test_size=0.2’ menunjukkan bahwa data *training* dan *testing* dibagi menjadi 80% dan 20%. Pembagian data tersebut menggunakan prinsip Pareto yaitu prinsip yang membagi data menjadi 80:20 (Loelianto et al., 2020). Data yang digunakan berjumlah 500 data sehingga pembagian yang dilakukan menjadi 400 untuk data *training* dan 100 untuk data *testing*. Kemudian, kode ‘random_state=42’ digunakan untuk melakukan pengacakan data yaitu data diacak sebanyak 42 kali.

3. Pembuatan Model

Pembuatan model merupakan langkah yang perlu dilakukan sebelum melakukan analisis. Pembuatan model pada penelitian ini menggunakan algoritma KNN. Berikut adalah kode yang digunakan untuk pembuatan model pada penelitian yang dilakukan.

```
#Model KNN
knn = KNeighborsClassifier()
knn.fit(X_train, y_train)
knn.score(X_train, y_train)
```

Kode baris pertama dari pemodelan yang dibuat berasal dari *library* sklearn. Kode baris kedua dari model digunakan untuk melatih model. Kemudian, kode baris ketiga dari model menunjukkan akurasi saat algoritma tersebut dilatih. Parameter ‘fit’ digunakan untuk menyesuaikan model dengan data pelatihan. Parameter ini

umumnya mengambil dua parameter utama yaitu nilai X dan y. Kemudian, evaluasi yang dilakukan menggunakan fungsi ‘score’ dapat memberikan gambaran dari performa model yang dilatih namun tetap perlu dilakukan pengujian untuk memastikan keseluruhan performa pembuatan model yang dilakukan dari algoritma. Score yang didapatkan dari algoritma KNN adalah 0.9525.

3.2.5 Tahap Analisis dan Hasil

Tahap analisis dan hasil. Analisis data untuk mengetahui kasus *stunting* dari bulan ke bulan setelah dilakukan pengolahan. Selain itu, juga untuk melihat prediksi jumlah balita antara yang terindikasi *stunting* dengan yang normal, mengetahui algoritma terbaik dari ketiga algoritma yang telah dilakukan pengujian, dan akurasi model yang didapatkan. Sedangkan hasil akhir dari penelitian ini yaitu berupa *dashboard* visualisasi data.

Dashboard visualisasi data terdiri dari beberapa halaman yaitu halaman dashboard yang digunakan sebagai menu utama dalam dashboard tersebut, halaman untuk mengupload data, dan halaman untuk klasifikasi data.