

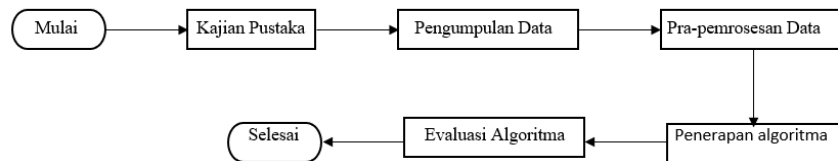
## BAB III METODE PENELITIAN

### 3.1 Objek Penelitian

Fokus utama penelitian ini yaitu kekerasan terhadap anak serta perempuan berdasarkan kelompok usia. Data yang digunakan pada penelitian ini tercatat dalam situs [opendata.jabarprov.go.id](http://opendata.jabarprov.go.id) mencakup data kekerasan yang terjadi pada tahun 2017 hingga 2023, di kabupaten dan kota Jawa Barat memiliki 18 kabupaten dan 9 kota. Terdiri dari 1323 data serta memiliki 9 atribut.

### 3.2 Prosedur Penelitian

Proses perancangan ini mencakup 5 proses yang diterapkan dengan bertahap. Proses penelitian ini mencakup kajian pustaka, pengumpulan data, pra-pemrosesan data, penerapan algoritma data *mining* serta evaluasi algoritma. Rangkaian proses tersebut dapat dilihat pada diagram alur yang tercantum dalam Gambar 3.1 berikut:



Gambar 3. 1 Urutan Proses Penelitian

#### 1. Kajian Pustaka

Dalam kajian pustaka, mempelajari berbagai sumber seperti jurnal dan skripsi yang membahas penerapan teknik data mining yang berfokus pada tahapan *klustering*, terutama Algoritma K-Means serta K-Medoids. Sumber-sumber informasi tersebut didapatkan melalui internet serta perpustakaan untuk memberikan landasan teori yang kuat pada penelitian ini.

#### 2. Pengumpulan Data

Proses pengumpulan data pada penelitian ini dilakukan melalui sumber situs [opendata.jabarprov.go.id](http://opendata.jabarprov.go.id). Adapun isi dataset tercantum dalam Tabel 3.1:

Tabel 3. 1 Data Korban Kekerasan Berdasarkan Kelompok Usia

Penelitian ini menggunakan data yang mencakup data kasus kekerasan pada

<b>Id</b>	<b>Kode Provinsi</b>	<b>Nama Provinsi</b>	<b>Kode kabupaten kota</b>	<b>Nama kabupaten kota</b>	<b>Kelompok Usia</b>	<b>Jumlah korban</b>	<b>Satuan</b>	<b>Tahun</b>
728	32	JAWA BARAT	3275	KOTA BEKASI	60+ TAHUN	2	ORANG	2020
480	32	JAWA BARAT	3215	KABUPATEN KARAWANG	18- 24 TAHUN	0	ORANG	2019
241	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	13-17 TAHUN	7	ORANG	2018
....				...				...
423	32	JAWA BARAT	3207	KABUPATEN CIAMIS	13-17 TAHUN	1	ORANG	2019
50	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	0-5 TAHUN	0	ORANG	2017
292	32	JAWA BARAT	3215	KABUPATEN KARAWANG	25-44 TAHUN	15	ORANG	2018
874	32	JAWA BARAT	3217	KABUPATEN BANDUNG BARAT	45- 59 TAHUN	1	ORANG	2021

anak serta perempuan di Jawa Barat pada rentang tahun 2017 hingga 2023, yang terdiri dari 1323 data dan 9 atribut.

### 3. Pra-pemrosesan Data

Proses berikutnya yaitu pra-pemrosesan data. Pra-pemrosesan data diterapkan melalui metode mengolah data mentah menjadi format yang lebih terstruktur agar sesuai sehingga siap digunakan pada tahap analisis selanjutnya. Tahapan ini mencakup seleksi data, pembersihan data, transformasi data, serta normalisasi data.

#### a. Seleksi Data

Tahap pertama dalam penelitian ini, melakukan seleksi data dengan menghapus beberapa kolom yang dianggap tidak relevan dari DataFrame.

Sebelum proses seleksi data memiliki 9 atribut dengan kolom id, kode provinsi, nama provinsi, kode kabupaten kota, nama kabupaten kota, kelompok usia, jumlah korban, satuan dan jumlah. Tabel 3.2 di bawah ini menyajikan data sebelum proses seleksi:

Tabel 3. 2 Sebelum Proses Seleksi Data

<b>Id</b>	<b>Kode Provinsi</b>	<b>Nama Provinsi</b>	<b>Kode kabupaten kota</b>	<b>Nama kabupaten kota</b>	<b>Kelompok Usia</b>	<b>Jumlah korban</b>	<b>Satuan</b>	<b>Tahun</b>
728	32	JAWA BARAT	3275	KOTA BEKASI	60+ TAHUN	2	ORANG	2020
480	32	JAWA BARAT	3215	KABUPATEN KARAWANG	18- 24 TAHUN	0	ORANG	2019
241	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	13-17 TAHUN	7	ORANG	2018
....				...				...
423	32	JAWA BARAT	3207	KABUPATEN CIAMIS	13-17 TAHUN	1	ORANG	2019
50	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	0-5 TAHUN	0	ORANG	2017
292	32	JAWA BARAT	3215	KABUPATEN KARAWANG	25-44 TAHUN	15	ORANG	2018
874	32	JAWA BARAT	3217	KABUPATEN BANDUNG BARAT	45- 59 TAHUN	1	ORANG	2021

Setelah proses seleksi data tersisa 3 atribut dengan kolom nama kabupaten kota, kelompok usia dan jumlah korban. Tabel 4.3 dibawah ini menyantumkan hasil sesudah proses seleksi data:

Tabel 3. 3 Sesudah Seleksi Data

<b>Nama Kabupaten Kota</b>	<b>Kelompok Usia</b>	<b>Jumlah Korban</b>
KABUPATEN KARAWANG	18-24 TAHUN	0
KABUPATEN KUNINGAN	13-17 TAHUN	7
KABUPATEN CIAMIS	13-17 TAHUN	1
....	....	
KABUPATEN KUNINGAN	0-5 TAHUN	0
KABUPATEN KARAWANG	25-44 TAHUN	15
KABUPATEN BANDUNG BARAT	45-59 TAHUN	1

Nama Kabupaten Kota	Kelompok Usia	Jumlah Korban
KABUPATEN KUNINGAN	0-5 TAHUN	3
KABUPATEN BOGOR	0-5 TAHUN	0

b. Pembersihan Data

Proses pembersihan data bertujuan untuk menghapus nilai yang kosong, mengatasi data yang duplikat serta menyaring atribut data yang tidak diperlukan. Pada Gambar 3.2 berikut ini menyajikan hasil dari tahapan pembersihan data:

```
Total Missing Values: 0
Jumlah Duplikasi Data: 0
Jumlah Duplikasi Data Setelah Menambahkan Noise: 0
Persentase Duplikasi Data: 0.00%
Persentase Missing Values: 0.00%
```

Gambar 3. 2 Pembersihan Data

Gambar 3.2 di atas terlihat bahwa tidak ada *missing value*, dan duplikasi data. Maka bisa untuk melanjutkan ke proses berikutnya.

c. Transformasi Data

Tahapan ini mencakup proses transformasi data yang diolah melalui cara mengelompokkan kelompok usia menjadi nilai numerik. Proses ini dilakukan dengan membuat *mapping* yang menghubungkan setiap rentang usia diberi kode angka untuk mempermudah analisis data lebih lanjut, seperti angka 1 untuk usia (0-5 tahun), angka 2 untuk usia (6-12 tahun), angka 3 untuk usia (13-17 tahun), angka 4 untuk usia (18-24 tahun), angka 5 untuk usia (25-44 tahun), angka 6 untuk usia (45-59 tahun) dan angka 7 untuk usia (60+ tahun). Berikut sebelum tahap transformasi data dapat dilihat pada Tabel 3.4:

Tabel 3. 4 Menampilkan Data Sebelum Proses Transformasi

Nama Kabupaten Kota	Kelompok Usia	Jumlah Korban
KOTA BEKASI	60+ TAHUN	2
KABUPATEN KARAWANG	18-24 TAHUN	0
KABUPATEN KUNINGAN	13-17 TAHUN	7

....	....	...
KABUPATEN CIAMIS	13-17 TAHUN	1
KABUPATEN	0-5 TAHUN	0
KUNINGAN		

Tabel 3.4 diatas yaitu data sebelum proses transformasi, mencakup tiga atribut yaitu nama kabupaten kota, kelompok usia dan jumlah korban. Dibawah ini Tabel 3.5 sesudah tahap transformasi data sebagai berikut:

Tabel 3. 5 Sesudah Data di Transformasi

Nama	1	2	3	4	5	6	7	Total
Kabupaten Kota								Korban
KABUPATEN BANDUNG	71	163	224	96	130	42	1	727
KABUPATEN BANDUNG BARAT	21	30	72	32	47	4	0	206
KABUPATEN BEKASI	75	134	192	118	236	35	3	793
....	....	....	....	....	....	....	....	....
KABUPATEN BOGOR	8	24	48	14	51	8	0	153
KABUPATEN CIAMIS	7	27	44	9	17	6	0	110

d. Normalisasi Data

Normalisasi data merupakan proses penyesuaian nilai variabel agar berada dalam skala yang setara. Proses ini bertujuan untuk mengubah nilai-nilai tersebut ke dalam rentang standar, yaitu antara 0 hingga 1, sehingga memudahkan dalam proses analisis dan perbandingan antar data. Berikut sebelum tahap normalisasi data dapat dilihat pada tabel 3.6:

Tabel 3. 6 Sebelum Normalisasi Data

Nama Kabupaten Kota	1	2	3	4	5	6	7	Total Korban
KABUPATEN BANDUNG	71	163	224	96	130	42	1	727
KABUPATEN BANDUNG BARAT	21	30	72	32	47	4	0	206
KABUPATEN BEKASI	75	134	192	118	236	35	3	793
...	....	...	...	....	....	...	....	...
KABUPATEN BOGOR	8	24	48	14	51	8	0	153
KABUPATEN CIAMIS	7	27	44	9	17	6	0	110

Pada proses normalisasi, seluruh kolom termasuk nama kabupaten kota ikut dalam proses transformasi sehingga nilai teksnya berubah menjadi angka. Namun, hal ini tidak mempengaruhi proses klastering karena nama kabupaten hanya digunakan sebagai identitas, bukan sebagai fitur analisis. Setelah proses normalisasi selesai, kolom nama kabupaten kota asli disisipkan kembali agar hasil pengelompokan dapat dikenali berdasarkan wilayah. Tahap normalisasi ini tercantum pada Tabel 3.7 berikut:

Tabel 3. 7 Sesudah Normalisasi Data

Nama Kabupaten Kota	1	2	3	4	5	6	7	Total Korban
0	0.503546	1.156028	1.588652	0.680851	0.921986	0.297872	0.007092	5.156028
1	0.148936	0.212766	0.51038	0.226950	0.033333	0.028369	0.000000	1460993
2	0.531915	0.950355	1.361702	0.836879	1.673759	0.248227	0.021277	5.624113
...	....	...	...	....	....	...	....	...
3	0.056738	0.170213	0.340426	0.099291	0.361702	0.056738	0.007092	1.085106
4	0.49644	0.191489	0.312057	0.063830	0.120567	0.042553	0.021277	0.780814

#### 4. Penerapan Algoritma Data Mining

Langkah awal data mining yaitu menentukan jumlah kluster dengan menggunakan metode *elbow*. Metode *elbow* ini bertujuan dalam penentuan nilai k pada

algoritma K-Means dan K-Medoids dengan merujuk pada grafik yang menampilkan nilai SSE (*Sum of Squared Errors*). Mengacu pada hasil jumlah kluster maka dilakukan penentuan pusat *centroid* atau kluster algoritma K-Means, setelah menentukan pusat *centroid*, kemudian menambahkan kolom kluster untuk menunjukkan hasil klustering. Pada algoritma K-Medoids, menentukan titik *centroid* berdasarkan data yang memiliki kedekatan paling tinggi dengan data lain dalam kluster. Setelah menentukan pusat kluster maka kolom tersebut ditambahkan untuk menandakan hasil pengelompokkan berdasarkan K-Medoids.

#### 5. Evaluasi Algoritma

Evaluasi terhadap hasil kluster dilakukan dengan menggunakan nilai *Silhouette Coefficient*. Metode ini digunakan untuk mengukur sejauh mana klusterisasi terbentuk dengan baik serta seberapa jelas pemisahan antar kluster. Nilai *Silhouette Coefficient* yang lebih besar menunjukkan performa klusterisasi yang lebih optimal.

