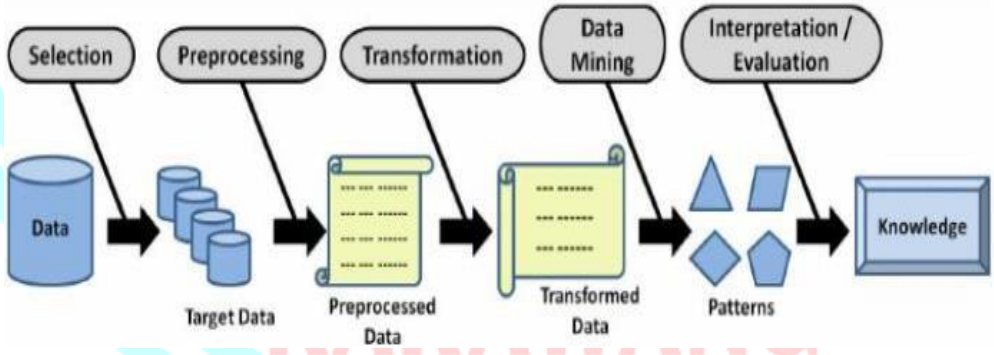
BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Data yang digunakan dalam penelitian ini diperoleh dari laman https://opendata.jabarprov.go.id/id/dataset. Data yang bersumber dari Dinas Pemberdayaan Perempuan, Perlindungan Anak dan Keluarga Berencana (DP3AKB) tersebut diakses pada 01 Juli 2024 pukul 20:30. Isinya yaitu mengenai data korban kekerasan berdasarkan jenis pekerjaan di Kabupaten/Kota Provinsi Jawa Barat dari tahun 2018-2022. Pengumpulan data yang diperoleh berjumlah 2160 data.

3.2 Prosedur Penelitian

Tahapan prosedur penelitian dapat dilihat pada gambar dibawah ini.



Gambar 3.1 Tahapan KDD

Penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD). pada gambar 3.1 dapat dilihat bahwa pada penelitian ini dilakukan secara berurutan dimulai dari seleksi data, *preprocessing* data, transformasi data, data mining, dan tahap terakhir yaitu evaluasi dari pengujian yang telah dilakukan.

3.3 Selection (Seleksi Data)

Seleksi data dilakukan karena data yang diperoleh tidak semuanya digunakan dalam penelitian ini. Seleksi data merupakan proses penyeleksian data dari atribut yang tidak diinginkan, ada beberapa atribut yang tidak relevan atau tidak dibutuhkan. Pada penelitian terdapat 3 atribut yang digunakan diantaranya yaitu

nama kabupaten atau kota, kategori pekerjaan dan jumlah korban. Dapat dilihat pada tabel 3.1.

Tabel 3.1 Seleksi Data

Nama Kabupaten Kota	Kategori Pekerjaan	Jumlah Korban
Kabupaten Bogor	Bekerja	0
Kabupaten Bogor	Ibu rumah tangga	0
Kabupaten Bogor	Tidakketahui	0
Kabupaten Bogor	Pedagang/tani/nelayan	0
Kabupaten Bogor	Pelajar	0
…	…	…
Kota Banjar	Pedagang/tani/nelayan	0
Kota Banjar	Pelajar	0
Kota Banjar	PNS/TNI/POLRI	0
Kota Banjar	Swasta/buruh	0
Kota Banjar	Tidak bekerja	0

3.4 Preprocessing (Pembersihan Data)

Pada tahap ini terdiri dari pemebersihan data untuk menghilangkan missing value dan duplikasi data.

3.5 Transformasi Data

Transformasi data digunakan untuk merubah bentuk suatu data dari sekumpulan data yang akan dibutuhkan dalam dataset. Data di transformasikan kedalam bentuk yang sudah ditambahkan dari tahun 2018-2022 untuk mempermudah pada proses pengolahan data.

3.6 Data Mining

Pada tahap data mining, metode yang digunakan dalam penelitian ini adalah *clustering*, yaitu teknik pengelompokan data yang dilakukan tanpa memerlukan label kelas (unsupervised learning). Tujuan dari clustering adalah untuk membagi data ke dalam sejumlah kelompok (klaster) berdasarkan tingkat kesamaan antar data, sehingga data dalam satu klaster memiliki karakteristik yang mirip dan berbeda dengan data pada klaster lain. Dalam proses ini, digunakan dua algoritma clustering, yaitu *K-Means* dan *K-Medoids*, yang bertujuan untuk melihat dan membandingkan hasil pengelompokan dari masing-masing metode. Berikut merupakan pseudocode untuk Algoritma K-Means (Cevin et al., n.d.).

Input:
Data
Jumlah klaster
max_iter: Jumlah iterasi maksimum (opsional)

Output:
Centroid
Klaster

Metode:

1. Inisialisasi:
   a. Pilih secara acak k centroid awal dari data.
   b. Tetapkan max_iter, jika perlu.

2. Ulangi langkah berikut sampai konvergensi atau mencapai max_iter:
   a. Assign Step
      Untuk setiap data point di dalam data:
      - Hitung jarak dari data point ke masing-masing centroid.
      - Tetapkan data kedalam klaster dengan centroid terdekat.
   b. Update Step
      Untuk setiap klaster:
      - Hitung centroid baru.
   c. Konvergensi
      - Jika centroid tidak berubah atau perubahan sangat kecil, hentikan iterasi.
3. Output:
   - Return centroid dari setiap klaster.
   - Return assignment data point ke masing-masing klaster.

Setelah melakukan perhitungan menggunakan *K-Means*, kemudian melakukan perhitungan mengunakan *K-Medoids*. Berikut adalah *pseudecode K-Medoids* (Nafilah et al., 2024)

Input:
Jumlah klaster
mean
Output:
c: nilai centroid
L: anggota
Metode:

1. Inisialisasi:
   a. Pilih secara acak k medoid awal dari data.
   b. Tetapkan max_iter, jika perlu.
2. Ulangi langkah berikut sampai konvergensi atau mencapai max_iter:
   a. Assign Step
      Untuk setiap data point di dalam Data:
      - Hitung jarak dari data point ke masing-masing medoid.
      - Tetapkan data point ke klaster dengan medoid terdekat.

   b. Update Step
      Untuk setiap cluster:
      - Pilih data point dalam klaster tersebut sebagai kandidat medoid baru.
      - Hitung total biaya (misalnya, jarak total antara semua titik di klaster dan medoid baru).
      - Jika medoid baru mengurangi total biaya, gantikan medoid lama dengan medoid baru.
   c. Konvergensi
      - Jika tidak ada perubahan pada medoid atau perubahan biaya sangat kecil, hentikan iterasi.
3. Output:
   - Return medoid dari setiap klaster.
   - Return assignment data point ke masing-masing klaster.

3.7 Interpretasi / Evaluasi

Pada penelitian ini, tujuan evaluasi dilakukan yaitu untuk melihat kualitas objek dalam suatu klaster yang telah dihasilkan dari perhitungan *K-Means* dan *K-Medoids* dengan menggunakan metode *Silhoutte Coefficient* dan *Davies Bouldin Index. Silhoutte Coefficient* merupakan kombinasi dari metode cohesion dan separation. Cohesion bertujuan untuk menghitung hubungan antara objek dalam klaster, Sementara itu, metode separation bertujuan untuk menilai seberapa jauh suatu klaster terpisah dari klaster lainnya. Dimana jumlah klaster terbaik adalah jumlah klaster yang memiliki rata-rata nilai silhouette coefficient tertinggi atau

mendekati 1. (Simanjuntak & Khaira, 2021). Berikut langkah-langkah perhitungan metode *silhouette coefficient* diantaranya adalah :

1. Menghitung jarak rata-rata dari suatu data misalkan i dengan semua data lain yang berada dalam satu klaster yang sama $a(i)$.

   $a(i) = \frac{1}{[A]-1} \sum j \in_{A,j \neq i} d(i,j)$

   Dengan j adalah data lain dalam satu klaster A dan d (i,j) adalah jarak antara data i dengan j.

2. Menghitung rata-rata jarak dari data i tersebut dengan semua data pada klaster yang lainnya dan ambil nilai terkecil.

   $d(i,C) = \frac{1}{[A]-1} \sum_j J \in C \; d(i,j)$

   Dengan $d(i,C)$ adalah jarak rata-rata data i dengan semua objek pada klaster lain C dimana A ≠ C.

   $b(i) = \min C \; \neq A d(i,C)/2$

3. Rumus *Silhoutte Coefficient* adalah :

   $$S(i) = \frac{b(i) - a(i)}{\max\,(a(i), b(i))/2}$$

   Dengan s(i) adalah semua rata-rata pada/2semua kumpulan data.

Davies Bouldin Index (DBI) merupakan metode evaluasi hasil *clustering*. Semakin kecil nilai DBI, maka semakin baik hasil klaster tersebut. Berikut adalah rumus *Davies Bouldin Index* (DBI) (Jollyta & Siddik, 2023).

$$DBI = \frac{1}{K} \sum_{j=1}^{K} \max_{j \neq k} Rj,k \qquad (5)$$

K merupakan klaster sedangkan R merupakan rasio perbandingan klaster.