

Klasterisasi Kasus Kekerasan Berdasarkan Jenis Lokasi Kejadian di Jawa Barat Menggunakan Algoritma *K-Means*

¹Cahya Diningrat, ²Bayu Priyatna, ³Elfina Novalia, ⁴Shofa Shofia Hilabi
^{1,2,3,4}Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang, Indonesia
si21.cahyadiningrat@mhs.ubpkarawang.ac.id, bayu.priyatna@ubpkarawang.ac.id,
elfinanovalia@ubpkarawang.ac.id, shofa.hilabi@ubpkarawang.ac.id

Submit : 05 Mei 2025 | Diterima : 12 Mei 2025 | Terbit : 13 Mei 2025

ABSTRAK

Kekerasan merupakan permasalahan sosial yang terjadi di berbagai lokasi dengan pola yang beragam. Di Provinsi Jawa Barat, kasus kekerasan menunjukkan keragaman pola yang dipengaruhi oleh faktor geografis, sosial, budaya, dan terutama jenis tempat kejadian. Memahami sebaran lokasi dan ciri-ciri kejadian kekerasan sangat penting untuk menyusun strategi penanganan yang tepat. Namun, kajian yang mengaitkan jenis tempat kejadian dengan frekuensi kekerasan di tingkat kabupaten dan kota masih terbatas. Oleh sebab itu, fokus utama dari penelitian ini adalah untuk mengelompokkan kasus kekerasan berdasarkan jenis lokasi kejadian menerapkan *Algoritma K-Means Clustering*. Proses penelitian mengikuti tahapan *Knowledge Discovery in Database (KDD)*, yang mencakup berbagai langkah, antara lain seleksi data, pra-pemrosesan data, transformasi data, *data mining*, serta interpretasi dan evaluasi. Hasil klasterisasi menunjukkan bahwa kasus kekerasan dapat dikelompokkan ke dalam tiga *cluster* utama berdasarkan jumlah kekerasan dan kategori jenis tempatnya dengan tingkatan kasus kekerasan yang berkategori tinggi, sedang, dan rendah. Evaluasi menggunakan *Silhouette Coefficient* menghasilkan nilai indeks 0,59, yang mengindikasikan bahwa kualitas *cluster* berada pada kategori cukup baik (*medium structure*). Dengan adanya hasil dari penelitian ini, diharapkan dapat menjadi sumber referensi untuk pihak-pihak yang berwenang dalam memahami pola kekerasan di berbagai lokasi, sehingga strategi pencegahan yang lebih efektif dapat dirancang untuk mengurangi tingkat kekerasan di Jawa Barat.

Kata Kunci: Kekerasan, Klasterisasi, *KDD*, *K-Means*, Jawa Barat

PENDAHULUAN

Kekerasan merupakan permasalahan sosial yang harus dihadapi dengan perhatian mendalam dari berbagai pihak terkait (Arief, Sudahri Damanik, & Irawan, 2021). Di Provinsi Jawa Barat, data menunjukkan bahwa kekerasan terjadi di berbagai lokasi dengan pola yang berbeda-beda. Selain itu, kebijakan daerah juga memainkan peran dalam tingkat prevalensi kekerasan di suatu wilayah. Oleh karena itu, diperlukan pendekatan analitis yang tepat untuk memahami karakteristik kejadian kekerasan dan pola persebarannya. Dengan pemahaman yang lebih baik, diharapkan strategi pencegahan dapat lebih efektif dan tepat sasaran dalam mengurangi tingkat kekerasan di masyarakat (Astuti & Muhammad Basysyar, 2024).

Selain faktor demografi dan sosial ekonomi, infrastruktur keamanan dan kebijakan daerah juga turut berperan dalam menentukan tingkat kekerasan di suatu wilayah. Daerah dengan sistem keamanan yang baik cenderung memiliki jumlah kasus kekerasan yang lebih rendah dibandingkan dengan daerah yang minim pengawasan. Oleh karena itu, pemetaan kasus berdasarkan lokasi dapat menjadi langkah awal dalam memahami kondisi sosial di suatu daerah.

Beberapa penelitian sebelumnya telah mengkaji performa algoritma *k-means* dalam mengelompokkan wilayah yang rawan terhadap kasus kekerasan pada Perempuan dan anak

berdasarkan usia. Pengujian dilakukan dengan menggunakan Rapidminer, dan validitas klusterisasi dievaluasi melalui *Davies-Bouldin Index (DBI)*, yang memiliki nilai 0,211 (Fauziah & Purnamasari, 2023). Dalam penelitian ini, jumlah *cluster* yang optimal ditentukan sebanyak 6 *cluster*. Di samping itu, metode *K-Means Clustering* juga dimanfaatkan dalam pengelompokan provinsi-provinsi di Indonesia berdasarkan data angka pengangguran terbuka. Hasil pengelompokan ini membagi provinsi ke dalam 3 *cluster*, dimana evaluasinya menggunakan *Silhouette Coefficient* dengan nilai 0,54, menandakan bahwa kategori pengelompokan berada dalam kategori *cluster* yang cukup baik (Nurfathullah & Purnamasari, 2024). Penelitian lain mengungkapkan bahwa penerapan algoritma *K-Means* dapat menghasilkan klusterisasi yang lebih unggul dibandingkan dengan algoritma *K-Medoids* ketika mengelompokan daerah yang rawan kriminalitas di Indonesia. Nilai *Davies-Bouldin Index (DBI)* tertinggi untuk *K-Means* adalah 0,463, dimana jumlah *cluster* $K=5$, sementara nilai *DBI* untuk *K-Medoids* tercatat sebesar 1,089 pada $K=2$ (Hoerunnisa, Dwilestari, Dikananda, Sunana, & Pratama, 2024).

Berdasarkan penelitian-penelitian sebelumnya, pada penelitian ini data dikelompokkan menggunakan metode *K-Means* berdasarkan kesamaan karakteristiknya, sehingga pola yang tersembunyi dalam data dapat diidentifikasi dengan lebih baik. Dengan menerapkan teknik ini, diharapkan dapat ditemukan hubungan antara jenis tempat kejadian dan frekuensi kekerasan yang terjadi. Pada penelitian ini, data kasus kekerasan yang ada di wilayah Jawa Barat akan dikelompokkan dengan menerapkan metode *K-Means*, dimana fokus pengelompokannya adalah berdasarkan jenis lokasi kejadian, Tahapan penelitian mengikuti pendekatan *Knowledge Discovery in Database (KDD)*, yaitu meliputi seleksi data, pra-pemrosesan data, transformasi data, *data mining*, serta interpretasi dan evaluasi (Maesaroh, Nur Padilah, & Haerul Jaman, 2023).

Diharapkan hasil dari penelitian ini dapat memberikan wawasan bagi pihak-pihak berwenang dalam menyusun kebijakan yang lebih efektif untuk pencegahan dan penanggulangan kasus kekerasan. Dengan mengetahui pola distribusi kekerasan berdasarkan lokasi kejadian, upaya mitigasi dapat dilakukan dengan lebih terarah sesuai dengan karakteristik masing-masing wilayah. Studi ini juga dapat dijadikan sebagai dasar dalam pengembangan kebijakan publik terkait pencegahan kekerasan di masa mendatang. pemerintah, dan organisasi sosial dapat menyusun program-program intervensi yang lebih sesuai dan tepat sasaran.

TINJAUAN PUSTAKA

Kekerasan

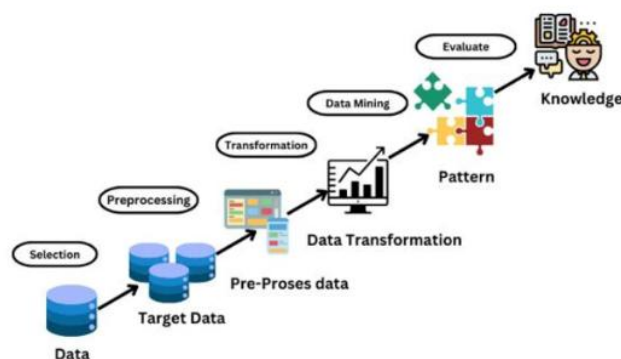
Kekerasan merupakan tindakan yang mengakibatkan kerugian fisik, psikologis, atau sosial terhadap individu atau kelompok (Rahma & Mufidah, 2022). Kekerasan dapat terjadi dimana saja, seperti di rumah tangga, tempat kerja, dan fasilitas umum (Noviya Adawiyah, 2021). Dalam konteks penelitian ini, analisis *cluster* digunakan untuk memahami pola kekerasan berdasarkan tempat kejadian.

Clustering

Clustering merupakan pendekatan statistik yang berfungsi dalam pengelompokan data atau menyusun objek ke dalam kelompok sesuai dengan aspek yang menjadi bagian dari data tersebut (Dwi Susilo, Shofiah Hilabi, Priyatna, & Novalia, 2024). Dalam pendekatan ini, algoritma mempelajari data berdasarkan kedekatannya, yang umumnya dikenal sebagai pengelompokan (Tyas Widyanti, 2023).

Data Mining

Data mining merupakan bidang kajian yang berorientasi pada pendekatan untuk memperoleh pemahaman atau mengenali pola dari data (Febianto & Palasara, 2019). *Data mining* juga dikenal sebagai *Knowledge Discovery in Database (KDD)* yang terdiri dari beberapa tahapan, yaitu seleksi data, pra-pemrosesan data, transformasi data, *data mining*, serta interpretasi dan evaluasi hasil (Helia & Lukman Rohmat, 2024).



Gambar 1 Tahapan Proses KDD

Algoritma K-Means

K-Means adalah sebuah algoritma yang mengelompokkan objek atau data menjadi beberapa bagian melalui proses partisi yang dilakukan secara iteratif, dengan maksud untuk memperkecil rata-rata jarak antara data dengan pusat cluster yang sesuai. Data dengan karakteristik serupa akan dikelompokkan ke dalam satu cluster, sementara data dengan karakteristik yang berbeda akan ditempatkan pada cluster lain (Baldah, Duarisah, & Maulana, 2023).

Berikut adalah rumus algoritma *K-Means*:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Keterangan:

$d(x, y)$ = Jarak *Euclidean* antara dua buah titik atau data x dan y

x = Vektor fitur/data pertama berdimensi- n .

y = Vektor fitur/data kedua berdimensi- n .

x_i, y_i = Nilai fitur ke- i dari vektor x dan y .

n = Jumlah dimensi atau jumlah fitur yang dimiliki oleh setiap data.

Metode Elbow

Metode *Elbow* merupakan teknik yang digunakan untuk menentukan jumlah (k) atau cluster yang tepat dalam dataset penelitian. Proses metode ini mengidentifikasi persentase hasil perbandingan dari total jumlah (k) dan juga menampilkan sebuah lekukan pada grafik, yang disebut sebagai siku pada titik tertentu. Titik siku ini menunjukkan jumlah *cluster* optimal, dimana penambahan *cluster* lebih lanjut tidak memberikan peningkatan yang signifikan dalam varians yang dijelaskan (Sulistiyawan, Hapsery, & Junita Ayu Arifahanum, 2021).

Python

Python merupakan bahasa pemrograman yang memungkinkan pelaksanaan instruksi secara langsung dan mengadopsi pendekatan berorientasi objek. Bahasa ini menerapkan sistem semantik dinamis yang meningkatkan keterbacaan sintaksnya. Selain itu, *python* juga mudah dipelajari karena dilengkapi dengan fitur manajemen memori otomatis (Manalu & Gunadi, 2022). *Python* merupakan jenis bahasa pemrograman yang mengutamakan kemudahan membaca kode, mengkombinasikan berbagai fitur dengan sintaks sederhana, dan juga dilengkapi banyak pustaka komprehensif. Di samping itu, *Python* telah menyediakan dukungan untuk beragam pendekatan pemrograman, termasuk pemrograman berorientasi objek serta imperatif.

Silhouette Coefficient

Silhouette Coefficient merupakan parameter yang dipakai dalam menilai kualitas pengelompokan pada suatu *cluster*. Nilainya berada dalam rentang -1 hingga 1, apabila semakin tinggi nilainya, semakin besar tingkat kemiripan antar objek dalam satu klaster dibandingkan dengan objek di luar klaster tersebut. di *cluster* yang berbeda (Atira & Nurina Sari, 2023).

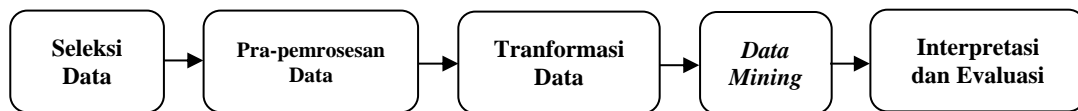
Google Colaboratory

Google Colab atau *Google Colaboratory* adalah sebuah *platform* pengembangan terbaru untuk bahasa pemrograman *python*. Pada *platform* ini, komputasi dilakukan menggunakan infrastruktur server *Google* yang dilengkapi dengan spesifikasi perangkat keras yang tinggi. *Google Colab* adalah aplikasi yang dikembangkan oleh *Google Research*. Program ini memfasilitasi pengguna untuk membuat dan mengeksekusi kode program melalui peramban web tanpa batasan, memberikan keuntungan signifikan dalam bidang machine learning, analisis data, dan pendidikan (Sitar & Leary, 2023).

METODE PENELITIAN

Metode Penelitian

Metodologi yang diterapkan dalam proses penelitian ini adalah dengan mengadopsi pendekatan *data mining* melalui prosedur penelitian *Knowledge Discovery in Database (KDD)*, dimana tujuannya adalah untuk mengekstraksi data yang sesuai dan membangun model yang dihasilkan oleh data yang kompleks (Arrosyad, Purnamasari, & Ali, 2024). Pendekatan *KDD* dalam penelitian ini menerapkan algoritma *K-Means* untuk mengelompokkan data, dan menjadikan data mining sebagai elemen kunci.



Gambar 2 Metode Penelitian

Berikut penjelasan setiap tahapan-tahapan *Knowledge Discovery in Database (KDD)* yang akan dilakukan pada penelitian:

1. Seleksi Data

Dilakukan pemilihan variabel yang relevan agar data yang digunakan dapat memberikan informasi yang akurat dan berguna. Pemilihan variabel ini merupakan bagian dari proses seleksi fitur yang bertujuan untuk menyaring atribut-atribut yang memiliki kontribusi signifikan terhadap tujuan analisis. Dengan demikian, proses ini juga dapat mengurangi kompleksitas dan waktu pemrosesan.

2. Pra-pemrosesan Data

Mengecek dataset apakah terdapat *missing value* atau tidak. *Missing value* dapat mengganggu analisis dan menghasilkan kesimpulan yang tidak akurat. Selain itu, keberadaan data yang hilang juga dapat mempengaruhi performa algoritma pemodelan, terutama pada metode yang sensitif terhadap nilai kosong. Oleh karena itu, penting untuk melakukan penanganan yang tepat terhadap *missing value*, seperti imputasi atau penghapusan, agar kualitas analisis tetap terjaga.

3. Transformasi Data

Mengubah data yang telah melalui data reprocessing agar dapat digunakan secara optimal oleh algoritma yang akan diterapkan. Proses transformasi data dalam penelitian ini mencakup inialisasi dataset, penyesuaian format data, serta penggantian nama variabel agar lebih sesuai untuk pemrosesan lebih lanjut di *Google Colab*.

4. Data Mining

Metode data mining yang diterapkan adalah pengelompokan menggunakan algoritma *K-Means* yang dijalankan dengan *Google Colab*. Algoritma *K-Means* digunakan untuk mengelompokkan data kasus kekerasan berdasarkan jenis tempat kejadian kekerasan. Metode *Elbow* digunakan untuk menentukan jumlah kluster yang paling optimal dengan menganalisis titik siku pada grafik.

5. Interpretasi dan Evaluasi

Silhouette Coefficient digunakan untuk mengevaluasi kualitas cluster yang dihasilkan dalam proses pengelompokan. Hasil dari evaluasi tersebut akan dijadikan dasar untuk memberikan rekomendasi yang berbasis data sehingga informasi yang dihasilkan lebih akurat dan aplikatif. Setelah proses evaluasi dilakukan, hasil klusterisasi kemudian divisualisasikan menggunakan *scatter plot* guna memberikan pemahaman visual terhadap pola pengelompokan yang terbentuk.

Sumber Data

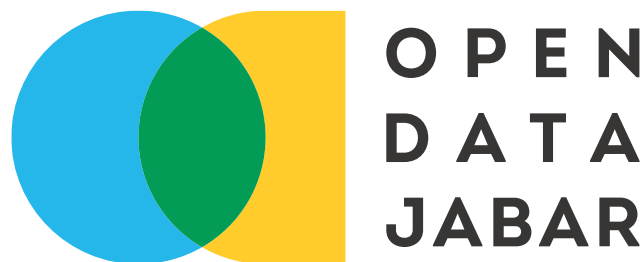
Dataset yang digunakan berasal dari *website* resmi Open Data Jabar yang meliputi berbagai kabupaten dan kota di Provinsi Jawa Barat. Dataset tersebut mencatat jumlah kasus kekerasan berdasarkan jenis lokasi kejadian selama periode 2018-2023, sehingga memungkinkan untuk melakukan analisis guna mengidentifikasi pola dan tren kejadian kekerasan. Sumber data ini berasal dari Dinas Pemberdayaan Perempuan, Perlindungan Anak, dan Keluarga Berencana, yang memiliki catatan resmi mengenai kasus kekerasan berdasarkan jenis lokasi kejadian. Dataset ini memiliki 972 data dengan 9 atribut, yaitu id, kode provinsi, nama provinsi, kode kabupaten dan kota, nama kabupaten dan kota, lokasi kejadian, jumlah kekerasan, satuan, dan tahun.

Tabel 1 Dataset Kasus Kekerasan di Jawa Barat

id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	tempat_kejadian	jumlah_kekerasan	satuan	tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	RUMAH TANGGA	3	KASUS	2018
2	32	JAWA BARAT	3201	KABUPATEN BOGOR	TEMPAT KERJA	0	KASUS	2018
3	32	JAWA BARAT	3201	KABUPATEN BOGOR	LAINNYA	2	KASUS	2018
4	32	JAWA BARAT	3201	KABUPATEN BOGOR	SEKOLAH	1	KASUS	2018
5	32	JAWA BARAT	3201	KABUPATEN BOGOR	FASILITAS UMUM	0	KASUS	2018
6	32	JAWA BARAT	3201	KABUPATEN BOGOR	LEMBAGA PENDIDIKAN KILAT	0	KASUS	2018
7	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	RUMAH TANGGA	8	KASUS	2018
8	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	TEMPAT KERJA	0	KASUS	2018
...
965	32	JAWA BARAT	3278	KOTA TASIKMALAYA	FASILITAS UMUM	0	KASUS	2023
966	32	JAWA BARAT	3278	KOTA TASIKMALAYA	LEMBAGA PENDIDIKAN KILAT	0	KASUS	2023
967	32	JAWA BARAT	3279	KOTA BANJAR	RUMAH TANGGA	1	KASUS	2023
968	32	JAWA BARAT	3279	KOTA BANJAR	TEMPAT KERJA	0	KASUS	2023
969	32	JAWA BARAT	3279	KOTA BANJAR	LAINNYA	4	KASUS	2023
970	32	JAWA BARAT	3279	KOTA BANJAR	SEKOLAH	0	KASUS	2023
971	32	JAWA BARAT	3279	KOTA BANJAR	FASILITAS UMUM	0	KASUS	2023
972	32	JAWA BARAT	3279	KOTA BANJAR	LEMBAGA PENDIDIKAN KILAT	0	KASUS	2023

Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan adalah dengan mengunduh dataset dari situs *web* Open Data Jabar, yang merupakan sumber yang dapat dipercaya. Selain itu, dilakukan proses verifikasi dan validasi data untuk memastikan akurasi. Verifikasi dilakukan dengan mencocokkan struktur dan isi data terhadap dokumentasi atau metadata yang tersedia, sedangkan validasi mencakup pengecekan terhadap konsistensi nilai, keutuhan entri, serta kesesuaian format data. Langkah ini penting untuk menjamin bahwa data yang digunakan dalam analisis bebas dari kesalahan teknis maupun konseptual, serta relevan dengan tujuan penelitian.



Gambar 3 Website Open Data Jabar

HASIL DAN PEMBAHASAN

Seleksi Data

Tahapan ini dimulai dengan pemilihan data yang relevan dari dataset yang berisi informasi kasus kekerasan yang terjadi di Jawa Barat berdasarkan lokasi kejadian dan tahun. Dalam proses ini, hanya atribut yang dianggap relevan untuk analisis yang diambil, yaitu nama kabupaten kota, tempat kejadian, jumlah kekerasan, dan tahun. Pemilihan ini bertujuan untuk memfokuskan analisis hanya pada informasi yang dibutuhkan dalam membentuk pola kekerasan berdasarkan lokasi dan jumlah kejadian.

	nama_kabupaten_kota	tempat_kejadian	jumlah_kekerasan	tahun
0	KABUPATEN BOGOR	RUMAH TANGGA	3	2018
1	KABUPATEN BOGOR	TEMPAT KERJA	0	2018
2	KABUPATEN BOGOR	LAINNYA	2	2018
3	KABUPATEN BOGOR	SEKOLAH	1	2018
4	KABUPATEN BOGOR	FASILITAS UMUM	0	2018
...
967	KOTA BANJAR	TEMPAT KERJA	0	2023
968	KOTA BANJAR	LAINNYA	4	2023
969	KOTA BANJAR	SEKOLAH	0	2023
970	KOTA BANJAR	FASILITAS UMUM	0	2023
971	KOTA BANJAR	LEMBAGA PENDIDIKAN KILAT	0	2023

972 rows x 4 columns

Gambar 4 Menampilkan Atribut Dataset yang Dipilih

Pra-pemrosesan Data

Setelah data diseleksi, tahap berikutnya adalah membersihkan dan mempersiapkan data agar siap untuk dianalisis. Dalam notebook ini, data diperiksa dari kemungkinan adanya nilai kosong (*missing values*) serta dilakukan transformasi untuk mengubah data kategorik menjadi numerik. Tahapan ini penting untuk memastikan bahwa algoritma *data mining* yang digunakan dapat bekerja secara optimal tanpa bias akibat skala atau format data yang tidak seragam.

```
Missing Values:
nama_kabupaten_kota    0
tempat_kejadian        0
jumlah_kekerasan       0
tahun                  0
dtype: int64
```

Gambar 5 Pengecekan *Missing Values* pada Dataset

Hasil dari pemeriksaan data yang diproses dataset menunjukkan tidak memiliki nilai kosong yang ditandai dengan angka 0. Hal ini menunjukkan bahwa semua informasi yang diperlukan tersedia secara lengkap.

Transformasi Data

Transformasi data dilakukan dengan menyusun atribut yang telah diproses menjadi bentuk vektor fitur yang akan digunakan dalam proses clustering. Pada tahap ini, hanya atribut numerik hasil normalisasi (*jumlah_kekerasan_scaled*) dan encoding (*tempat_kejadian_encoded*) yang digunakan. Tahapan ini memungkinkan data direpresentasikan dalam ruang fitur yang sesuai untuk analisis kluster dengan metode K-Means.

	tempat_kejadian	tempat_kejadian_encoded	jumlah_kekerasan	jumlah_kekerasan_scaled
0	RUMAH TANGGA	3	3	-0.249851
1	TEMPAT KERJA	6	0	-0.378915
2	LAINNYA	1	2	-0.292872
3	SEKOLAH	4	1	-0.335894
4	FASILITAS UMUM	0	0	-0.378915
...
967	TEMPAT KERJA	6	0	-0.378915
968	LAINNYA	1	4	-0.206830
969	SEKOLAH	4	0	-0.378915
970	FASILITAS UMUM	0	0	-0.378915
971	LEMBAGA PENDIDIKAN KILAT	2	0	-0.378915

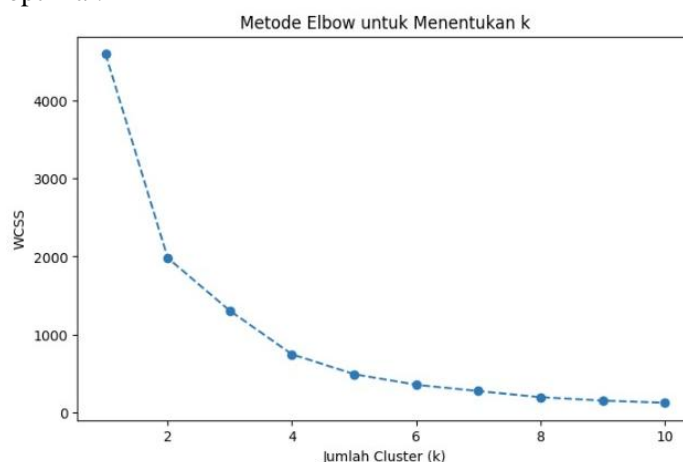
972 rows x 4 columns

Gambar 6 Inisialisasi Atribut pada Dataset

Data Mining

Tahap inti dari metode *KDD* adalah *data mining*, di mana dilakukan penerapan jenis algoritma *clustering* untuk menemukan pola tersembunyi dalam data. Dalam hal ini, digunakan algoritma *K-Means*. Tujuan dari tahap ini adalah untuk mengelompokkan data ke dalam kluster berdasarkan kesamaan karakteristik jumlah kekerasan dan lokasi kejadian. Hasil klusterisasi dapat memberikan wawasan mengenai wilayah atau jenis tempat kejadian dengan tingkat kekerasan yang serupa.

Pada tahap ini, metode *Elbow* digunakan untuk mengidentifikasi jumlah kluster yang paling optimal dengan menganalisis titik siku pada grafik. Metode ini bekerja dengan cara menghitung total jarak kuadrat antara data dengan pusat kluster (*centroid*) untuk berbagai jumlah *cluster*. Seiring bertambahnya jumlah kluster, nilai total jarak kuadrat akan berkurang. Titik siku pada grafik, yaitu titik di mana penurunan nilai total jarak kuadrat mulai melambat, menunjukkan jumlah kluster yang paling optimal.



Gambar 7 Menampilkan Grafik Metode *Elbow*

Langkah selanjutnya adalah menerapkan algoritma *K-Means* dengan jumlah kluster yang telah ditentukan berdasarkan hasil analisis menggunakan metode *Elbow*, di mana titik optimalnya diperoleh pada nilai $K=3$. Setelah proses klusterisasi dilakukan, kolom baru yang merepresentasikan hasil kluster (atribut *cluster*) ditambahkan ke dalam dataset sebagai informasi tambahan untuk setiap entri data.

	nama_kabupaten_kota	tempat_kejadian	jumlah_kekerasan	cluster
0	KABUPATEN BOGOR	RUMAH TANGGA	3	1
1	KABUPATEN BOGOR	TEMPAT KERJA	0	1
2	KABUPATEN BOGOR	LAINNYA	2	0
3	KABUPATEN BOGOR	SEKOLAH	1	1
4	KABUPATEN BOGOR	FASILITAS UMUM	0	0
...
967	KOTA BANJAR	TEMPAT KERJA	0	1
968	KOTA BANJAR	LAINNYA	4	0
969	KOTA BANJAR	SEKOLAH	0	1
970	KOTA BANJAR	FASILITAS UMUM	0	0
971	KOTA BANJAR	LEMBAGA PENDIDIKAN KILAT	0	0

972 rows × 4 columns

Gambar 8 Pengelompokan Data menggunakan Algoritma *K-Means*

Interpretasi dan Evaluasi

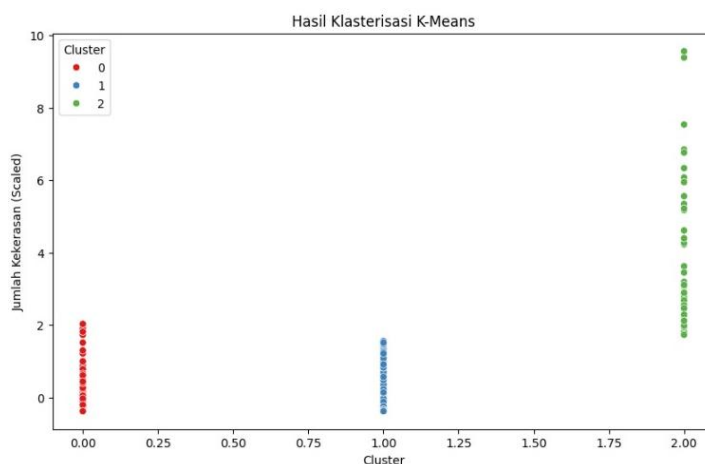
Tahapan terakhir adalah mengevaluasi dan menginterpretasi hasil dari proses *data mining*. Evaluasi dilakukan dengan menghitung nilai *Silhouette Score* yang mengukur seberapa baik objek-objek dalam satu *cluster* serupa satu sama lain, dan berbeda dengan *cluster* lainnya.

```
# 5. Evaluasi Pola (Pattern Evaluation)
# Evaluasi Hasil Clustering dengan Silhouette Score
sil_score = silhouette_score(df[['tempat_kejadian_encoded', 'jumlah_kekerasan_scaled']], df['cluster'])
print(f"Silhouette Score: {sil_score}")
```

Silhouette Score: 0.5945668431621698

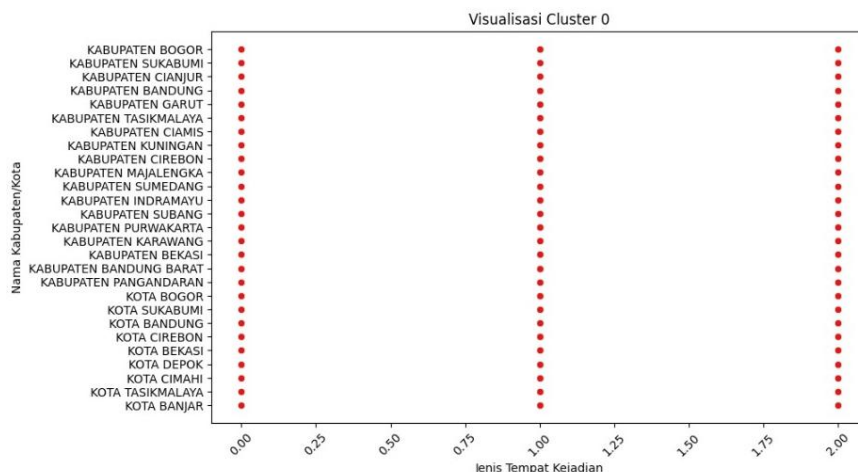
Gambar 9 Hasil Evaluasi menggunakan *Silhouette Coefficient*

Berdasarkan gambar 9, indeks nilai yang dihasilkan dengan menggunakan *Silhouette Coefficient* adalah 0,59. Hasil yang diperoleh mengindikasikan bahwa kualitas *cluster* berada pada kategori yang cukup baik (*medium structure*). Selanjutnya, hasil klasterisasi akan divisualisasikan dengan menggunakan *scatter plot*.



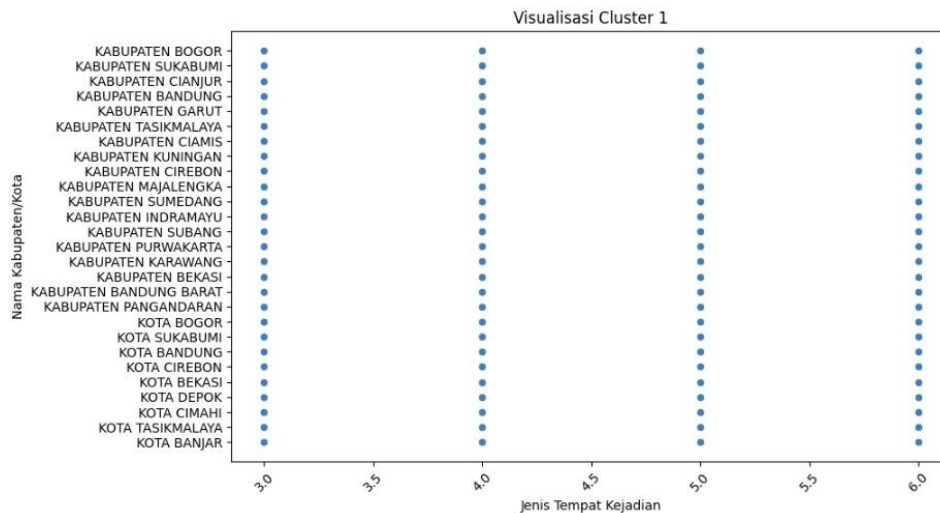
Gambar 10 Visualisasi Data

Berdasarkan Gambar 10, terlihat distribusi kasus kekerasan di Jawa Barat yang dikelompokkan ke dalam tiga *cluster*. *Cluster 0* ditandai dengan warna merah merupakan *cluster* dengan kategori tinggi, *Cluster 1* yang berwarna biru adalah *cluster* dengan kategori sedang, dan *Cluster 2* yang berwarna hijau adalah *cluster* dengan kategori rendah. Untuk penjelasan yang lebih mendetail, dapat merujuk pada visualisasi masing-masing *cluster* berikut.



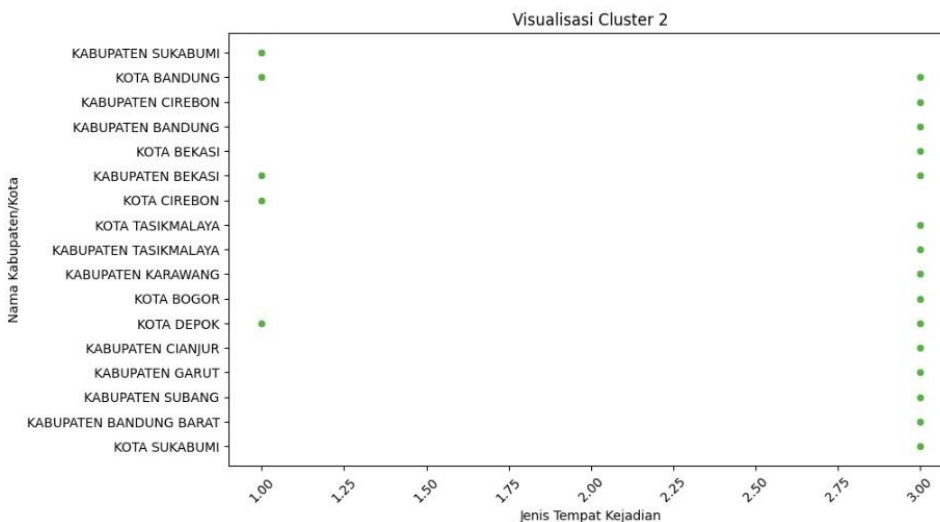
Gambar 11 Visualisasi Data *Cluster 0*

Berdasarkan Gambar 11, kasus kekerasan dengan kategori tinggi yang terjadi di lokasi seperti rumah tangga, sekolah, dan tempat kerja ditemukan di seluruh wilayah, mencakup 27 kabupaten/kota. Penjelasan mengenai kasus kekerasan dengan kategori sedang akan disampaikan pada Gambar 12.



Gambar 12 Visualisasi Data Cluster 1

Berdasarkan Gambar 12, kasus kekerasan dengan kategori sedang terjadi di lokasi seperti rumah tangga, sekolah, lembaga pendidikan kilat, dan tempat kerja yang tersebar di seluruh wilayah, mencakup 27 kabupaten/kota. Penjelasan mengenai kasus kekerasan dengan kategori rendah akan disampaikan pada Gambar 13.



Gambar 13 Visualisasi Data Cluster 2

Berdasarkan Gambar 13, kasus kekerasan dengan kategori rendah yang terjadi di lokasi lainnya ditemukan di 5 wilayah, yaitu Kota Depok, Kota Cirebon, Kabupaten Bekasi, Kota Bandung, dan Kota Sukabumi. Sementara itu, untuk jenis tempat kejadian rumah tangga, kasus tersebut terdapat di 15 wilayah, antara lain Kota Sukabumi, Kabupaten Bandung Barat, Kabupaten Subang, Kabupaten Garut, Kabupaten Cianjur, Kota Depok, Kota Bogor, Kabupaten Karawang, Kabupaten Tasikmalaya, dan Kota Tasikmalaya. dengan sebaran wilayah yang relatif kecil. Hal ini dapat mencerminkan efektivitas intervensi kebijakan lokal, karakteristik sosial budaya masyarakat setempat, atau kemungkinan rendahnya pelaporan kasus.

KESIMPULAN

Hasil penelitian menunjukkan bahwa kasus kekerasan dapat dikelompokkan menjadi tiga *cluster*, antara lain tinggi, sedang, dan rendah. Kasus berkategori tinggi dan sedang tersebar di seluruh 27 kabupaten/kota, dengan lokasi kejadian bervariasi seperti rumah tangga, sekolah, tempat kerja, fasilitas umum, dan lembaga pendidikan kilat. Sementara itu, kasus berkategori rendah terbatas pada 5 wilayah untuk jenis lokasi kejadian lainnya dan 15 wilayah untuk jenis tempat kejadian rumah tangga. Hasil evaluasi dengan *Silhouette Coefficient* menunjukkan kualitas cluster yang cukup baik dimana nilai yang dihasilkan sebesar 0,59. Oleh karena itu, fokus perhatian dan upaya penanggulangan kekerasan perlu diarahkan pada wilayah dengan kategori tinggi dan sedang untuk menekan jumlah kasus kekerasan yang terjadi di masyarakat.

REFERENSI

- Arief, N., Sudahri Damanik, I., & Irawan, E. (2021). KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Algoritma K-Medoids Dalam Mengelompokkan Tingkat Kasus Kejahatan di Setiap Provinsi. *Media Online*, 2(3), 111–116. Retrieved from <https://djournal.com/klik>
- Arrosyad, A. A., Purnamasari, A. I., & Ali, I. (2024). IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS PERSEBARAN UMKM DI JAWA BARAT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8).
- Astuti, R., & Muhammad Basysyar, F. (2024). PENERAPAN DATA MINING CLUSTERING MENGGUNAKAN METODE K-MEANS PADA DATA TINDAK KRIMINALITAS DI POLRES KABUPATEN KUNINGAN. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8).
- Atira, A., & Nurina Sari, B. (2023). Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan. *Jurnal Ilmiah Wahana Pendidikan*, 9(17), 76–86. <https://doi.org/10.5281/zenodo.8282638>
- Baldah, A., Duarisah, V., & Maulana, R. A. (2023). *Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode K-Means*. Retrieved from <https://www.bps.go.id>
- Dwi Susilo, D., Shofiah Hilabi, S., Priyatna, B., & Novalia, E. (2024). *Implementasi Data Mining dalam Pengelompokan Data Pembelian Menggunakan Algoritma K-Means Pada PT.Otomotif 1*.
- Fauziah, R., & Purnamasari, A. I. (2023). Implementasi Algoritma K-Means pada Kasus Kekerasan Anak dan Perempuan Berdasarkan Usia. *Hello World Jurnal Ilmu Komputer*, 2(1), 34–41. <https://doi.org/10.56211/helloworld.v2i1.232>
- Febianto, N. I., & Palasara, N. (2019). Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 8(2), 130–140. <https://doi.org/10.32736/sisfokom.v8i2.653>
- Helia, A., & Lukman Rohmat, C. (2024). ANALISIS PENGELOMPOKAN DAYA TARIK OBYEK WISATA BERDASARKAN JENISNYA MENGGUNAKAN METODE K-MEANS PADA DATA PEMPROV JABAR. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8).
- Hoerunnisa, A., Dwilestari, G., Dikananda, F., Sunana, H., & Pratama, D. (2024). KOMPARASI ALGORITMA K-MEANS DAN K-MEDOIDS DALAM ANALISIS PENGELOMPOKAN DAERAH RAWAN KRIMINALITAS DI INDONESIA. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8). Retrieved from <https://opendata.jabarprov.go.id/id>

-
- Maesaroh, M., Nur Padilah, T., & Haerul Jaman, J. (2023). PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA PENGELOMPOKAN DAERAH PENYEBARARAN DIARE DI PROVINSI JAWA BARAT. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7).
- Manalu, D. A., & Gunadi, G. (2022). IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI. *Infotech: Journal of Technology Information*, 8(1), 43–54. <https://doi.org/10.37365/jti.v8i1.131>
- Noviya Adawiyah, N. S. and ohamad J. (2021). KLASTERISASI KASUS KEKERASAN TERHADAP ANAK DAN PEREMPUAN BERDASARKAN ALGORITMA K-MEANS. *Generation Journal*, 5.
- Nurfathullah, M., & Purnamasari, I. (2024). IMPLEMENTASI K-MEANS UNTUK MENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDEKS JUMLAH PENGANGGURAN TERBUKA. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8).
- Rahma, R., & Mufidah, R. (2022). *PENGELOMPOKAN DAERAH RAWAN KEKERASAN TERHADAP PEREMPUAN DAN ANAK DI JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS*.
- Sitar, M. C., & Leary, R. J. (2023). Technical note: Colab_zirc_dims: A Google Colab-compatible toolset for automated and semi-Automated measurement of mineral grains in laser ablation-inductively coupled plasma-mass spectrometry images using deep learning models. *Geochronology*, 5(1), 109–126. <https://doi.org/10.5194/gchron-5-109-2023>
- Sulistiyawan, E., Hapsery, A., & Junita Ayu Arifahanum, L. A. (2021). *PERBANDINGAN METODE OPTIMASI UNTUK PENGELOMPOKAN PROVINSI BERDASARKAN SEKTOR PERIKANAN DI INDONESIA (Studi Kasus Dinas Kelautan dan Perikanan Indonesia)*. 10(1), 76–84. Retrieved from <https://ejournal3.undip.ac.id/index.php/gaussian/>
- Tyas Widyanti, S. S. H. A. H. T. E. N. (2023). Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi. *Jurnal Informasi Dan Teknologi*.