

IMPLEMENTASI ALGORITMA K-MEANS DALAM ANALISIS CLUSTER KORBAN KEKERASAN DI PROVINSI JAWA BARAT

Lia Awaliyah¹, Nining Rahaningsih², Raditya Danar Dana³

¹ Program Studi Teknik Informatika, STMIK IKMI Cirebon

² Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon

³ Program Studi Manajemen Informatika, STMIK IKMI Cirebon

Jalan Perjuangan No. 10B Majasem, Kec. Kesambi, Kota Cirebon, Tlp. 0231-490480-480481

liaawaliyah498@gmail.com

ABSTRAK

Kekerasan diartikan sebagai tindakan yang menyebabkan kerugian fisik, psikologis, atau perampasan hak korban. Pada tahun 2020-2022, tercatat peningkatan kasus kekerasan di Indonesia, Provinsi Jawa Barat menjadi salah satu yang paling terdampak. Kurangnya pemahaman mengenai pola dan karakteristik korban menjadi hambatan upaya perlindungan. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan korban kekerasan berdasarkan kategori pendidikan, jenis kelamin, jumlah dan Kabupaten/Kota dengan memanfaatkan metode algoritma *k-means clustering*. Hasil pengujian menggunakan tools *Rapidminer* 10.2 dan metrik evaluasi *DBI*, menunjukkan *cluster* optimal pada $K=2$, dengan nilai validitas 0,157. *Cluster 0* mencakup Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Karawang, Kabupaten Pangandaran, Kabupaten Purwakarta, Kota Bandung, Kota Banjar, Kota Bekasi, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Depok, Kota Sukabumi, dan Kota Tasikmalaya, dimana korban laki-laki didominasi tingkat pendidikan NA, tidak sekolah, SD, SLTA, dan TK dan korban perempuan didominasi tingkat pendidikan NA, SD, SLTA, Perguruan Tinggi, TK, dan PAUD. *Cluster 1* mencakup Kabupaten Bandung, Kabupaten Bogor, Kabupaten Ciamis, Kabupaten Cianjur, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Kuningan, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Sukabumi, Kabupaten Sumedang, dan Kabupaten Tasikmalaya, dimana korban laki-laki didominasi tingkat pendidikan SLTP dan PAUD dan korban perempuan didominasi tingkat pendidikan tidak sekolah dan SLTP.

Kata Kunci : Kekerasan, Clustering, K-means .

1. PENDAHULUAN

Kekerasan diartikan sebagai perbuatan atau ancaman terhadap diri sendiri, sekelompok orang, atau masyarakat yang menyebabkan kematian, kerusakan psikologis, trauma, atau pelanggaran hak. [1]. Tindakan fisik, baik di sengaja atau tidak disengaja, atau tindakan lain yang dapat berdampak negatif pada fisik, psikologis dan hak individu, dianggap sebagai kekerasan[2]. Dampak dari tindak kekerasan tersebut dianggap sangat luas dan berlangsung dalam jangka waktu yang lama. Trauma yang dialami oleh korban kekerasan akan diingat sepanjang hidupnya, terutama korban yang menerima tindak kekerasan merupakan anak-anak. Korban yang mendapatkan tindak kekerasan dari semasa kecil dianggap memiliki potensi untuk menjadi pelaku kekerasan dikemudian hari. Oleh karena itu, trauma yang dialami oleh korban kekerasan ini dianggap perlu mendapatkan perhatian khusus dan perlindungan yang efektif[3].

Pada tahun 2020, tercatat 20.499 kasus kekerasan di Indonesia. Angka tersebut meningkat pada tahun 2021 menjadi 25.210 kasus, dan terus mengalami peningkatan pada tahun 2022, jumlah kasus kekerasan mencapai 27.593. Salah satu provinsi di Indonesia yang mencatat jumlah kekerasan tertinggi adalah Provinsi Jawa Barat selama kurun waktu 2021-2022, setelah Jawa Timur dan Jawa Tengah[4]. Masalah kekerasan ini semakin

diperparah oleh kurangnya pemahaman mengenai pola dan karakteristik korban kekerasan di Provinsi Jawa Barat. Hal ini mengakibatkan upaya perlindungan dan pencegahan yang kurang efektif karena minimnya wawasan mengenai pola dan karakteristik korban.

Penelitian terdahulu yang membahas kinerja algoritma *k-means* dalam mengkategorikan daerah yang rentan kekerasan terhadap perempuan dan anak di Jawa Barat, hasil pengelompokan membagi daerah tersebut menjadi 3 klaster, dengan evaluasi menggunakan *Silhouette Coefficient* menghasilkan nilai 0,62, menunjukkan kriteria pengelompokan tergolong dalam struktur klaster standar[2]. Metode *k-means clustering* juga dimanfaatkan untuk mengkaji pola penerimaan beasiswa dari Bank Indonesia di Provinsi Riau. Hasil pengujian dengan *Rapidminer* dan evaluasi menggunakan *Davies Bouldin Index (BDI)* validitas sebesar 0,121, dengan optimalisasi jumlah kelompok terbaik mencapai 5 *cluster*[5]. Penelitian lain menyimpulkan bahwa algoritma *k-means* menunjukan kinerja yang baik dalam mengelompokkan bencana alam tanah longsor tahun 2019 di provinsi Jawa Barat, dibanding algoritma K-Medoids. Uji validitas *DBI* algoritma *k-means* sebesar 0,265 terdapat pada $K=6$, sedangkan nilai *DBI* algoritma *k-medoids* sebesar 0,342 terdapat pada $K=2$ [6]. Penerapan *DBI* sebagai metrik evaluasi juga digunakan untuk mengevaluasi seberapa baik

pengelompokan terbentuk dengan membagi *cluster* berdasarkan kesamaan dalam *cluster* dan perbedaan antar *cluster* pada data jumlah desa yang memiliki fasilitas pendidikan di Indonesia. Hasil penelitian menunjukkan bahwa *DBI* optimal diperoleh ketika jumlah *cluster* *K* adalah 2, dengan nilai *DBI* sebesar 0,168 untuk tipe pengukuran *Mixed Measures*[7].

Berdasarkan landasan tersebut, penelitian ini bertujuan untuk menganalisis dan mengelompokkan korban kekerasan di Provinsi Jawa Barat dengan menerapkan algoritma *k-means*. Penelitian juga bertujuan memberikan wawasan mengenai pola dan karakteristik korban kekerasan berdasarkan jumlah *cluster* optimal yang terbentuk melalui pengujian *cluster* model menggunakan metrik evaluasi *Davies Bouldin Index (BDI)*.

Penelitian ini mengimplementasikan algoritma *k-means*, karena algoritma *k-means* telah terbukti efektif dalam melakukan pengelompokan data tanpa supervisi (*unsupervised learning*) berdasarkan atribut-atribut yang dimiliki oleh data tersebut[8]. Algoritma *k-means* cenderung lebih efisien saat memproses sejumlah besar objek, karena akurasi yang optimal terhadap ukuran objek[2]. Oleh karena itu, algoritma *k-means* dipilih sebagai metode pengelompokan dalam penelitian ini untuk menentukan jumlah *cluster* optimal dan mengidentifikasi pola dan karakteristik dalam dataset jumlah korban kekerasan berdasarkan kategori pendidikan, jenis kelamin, jumlah korban, dan kabupaten/kota di Provinsi Jawa Barat.

2. TINJAUAN PUSTAKA

2.1. Kekerasan

Kekerasan dalam pengertian yang sempit merujuk pada tindakan seperti penyerangan, perusakan, atau penghancuran fisik diri sendiri atau sesuatu yang dapat menjadi milik orang lain. Sementara itu, dalam arti yang lebih luas, kekerasan mencakup tindakan psikologis atau fisik yang dilakukan oleh individu atau kelompok dengan sengaja atau tidak sengaja, baik secara eksklusif maupun tidak eksklusif[9].

2.2. Data Mining

Data Mining merupakan proses di mana korelasi-korelasi dan pola-pola yang baru dan bermanfaat ditemukan melalui eksplorasi dalam repositori data yang besar. Teknologi pengenalan pola seperti teknik statistik dan matematika digunakan pada proses ini. *Data mining* menggunakan teknik pembelajaran mesin untuk secara otomatis menganalisis dan mengekstrak pengetahuan[10].

2.3. Knowledge Discovery in Database

Knowledge Discovery in Databases (KDD) merupakan rangkaian proses yang melibatkan pengumpulan dan pemanfaatan data historis, dengan

tujuan mengidentifikasi pola atau keterkaitan dalam kumpulan data yang besar[11].

2.4. Clustering

Clustering merupakan metode statistik yang diterapkan untuk mengelompokkan sejumlah data atau objek besar ke dalam klaster berdasarkan karakteristik yang dimiliki oleh data atau objek tersebut. *Clustering* merupakan metode *unsupervised* dimana karakteristik dari setiap kelompok tidak ditentukan sebelumnya, tetapi justru ditentukan berdasarkan kesamaan atribut di dalam kelompok tersebut[12]. *Unsupervised learning* merupakan salah satu metode algoritma pembelajaran mesin yang digunakan untuk membuat kesimpulan dari dataset. Dalam metode ini, algoritma hanya belajar dari data berdasarkan kedekatannya atau sering disebut sebagai pengelompokan[13].

2.5. K-means

K-means merupakan metode pendekatan yang mempartisi data ke dalam kelompok melalui serangkaian partisi berulang, dengan tujuan mengurangi jarak rata-rata antar setiap data dan pusat klaster yang sesuai[14]. Data dengan karakteristik serupa akan dikelompokkan ke dalam satu klaster, sementara data dengan karakteristik yang berbeda akan ditempatkan dalam klaster lainnya[15]. Berikut adalah langkah-langkah pengelompokan menggunakan algoritma *k-means*:

1. Tentukan jumlah *K (cluster)* yang akan dibentuk
2. Tentukan titik awal untuk setiap *cluster*.
3. Hitung jarak setiap data ke pusat *cluster* dengan rumus jarak *Euclidean Distance* hingga menemukan jarak terdekat setiap data ke pusat klaster. Berikut adalah persamaan *Euclidean Distance*:

$$D(x, y) = \sqrt{(x_2 - y_1)^2 + (x^2 - y^2)^2} \dots\dots\dots(1)$$

Keterangan:

D menunjukkan jarak *Euclidean*.

(*x*₁, *y*₁) menunjukkan koordinat titik awal.

(*x*₂, *y*₂) menunjukkan koordinat titik kedua.

4. Kelompokkan data berdasarkan kedekatannya dengan *centroid*.
5. Hitung ulang *centroid* dengan nilai rata-rata dari semua objek data dalam *cluster* tersebut.
6. Hitung ulang setiap objek menggunakan *centroid*. Jika pusat *cluster* tidak berubah, maka proses pengelompokan selesai. Kembali ke proses 3 jika pusat *cluster* masih berubah.

2.6. Euclidean Distance

Dalam *k-means clustering*, *Euclidean Distance* digunakan untuk mengukur seberapa dekat atau jauh jarak suatu titik data dari *centroid* suatu *cluster*. Kelebihan *Euclidean Distance* terletak pada perannya sebagai metrik jarak yang umum digunakan dalam algoritma *k-means*, memberikan hasil yang lebih optimal dibandingkan dengan metode perhitungan lain. Namun, perhitungan manual *Euclidean Distance*

merupakan suatu kelemahan karena melibatkan jumlah iterasi yang lebih besar dan memerlukan banyak waktu[16].

2.7. DBI

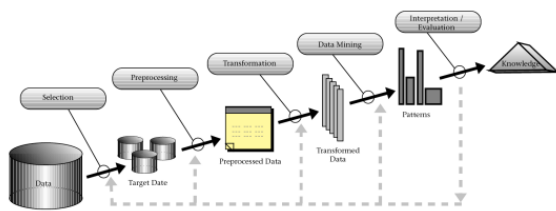
Indeks *Davies-Bouldin (DBI)* merupakan metode validasi untuk mengukur kualitas hasil pengelompokan. Sebuah kluster dianggap optimal apabila memiliki nilai *DBI* yang semakin rendah[9].

2.8. Rapidminer

Rapidminer merupakan aplikasi yang berperan sebagai alat pembelajaran dalam bidang *Data Mining*. Platform ini dikembangkan oleh perusahaan yang mengkhususkan diri dalam berbagai tahapan data besar untuk kebutuhan bisnis, penelitian, pendidikan, dan pembelajaran. *Rapidminer* menyediakan berbagai solusi pembelajaran untuk pengelompokan, klasifikasi, asosiasi, dan analisis regresi[17].

3. METODE PENELITIAN

Penelitian ini menerapkan metode *Data Mining* dengan pendekatan *Knowledge Discovery in Database(KDD)*.



Gambar 1. Alur tahap KDD
Sumber: infovis-wiki.net

Berikut rincian tahapan *KDD* yang dilakukan dalam penelitian:

1. **Data**
Identifikasi dataset dengan memilih dan menentukan data yang akan digunakan, yang mencakup informasi korban kekerasan.
2. **Selection**
Memilih dataset yang akan digunakan serta menyeleksi atribut-atribut yang dianggap paling relevan untuk tujuan analisis.
3. **Preprocessing**
Melakukan pembersihan data dan mengatasi nilai-nilai yang hilang dalam dataset.
4. **Transformation**
Melakukan transformasi bentuk atau tipe pada data, agar sesuai dengan algoritma yang akan dipakai saat pengolahan data pada tahap *Data Mining*.
5. **Data Mining**
Melakukan pengujian dengan menerapkan algoritma *K-means* dan tools *Rapidminer* untuk membentuk model *cluster*.

6. Evaluation

Mengevaluasi hasil pembentukan model *cluster* dengan metrik evaluasi *DBI* untuk mengukur kualitas model *cluster* yang terbentuk.

3.1. Pengumpulan Data

1. Studi literatur

Studi literatur melibatkan tinjauan literatur terhadap penelitian-penelitian sebelumnya. Prndrkatan ini mencakup pengumpulan, analisis, evaluasi, serta pengorganisasian informasi dari beragam sumber, seperti buku, artikel, jurnal, dan dokumen lain yang relevan dengan topik atau permasalahan penelitian. Tujuan dari *studi literatur* adalah untuk mendapatkan pemahaman mengenai topik atau masalah penelitian, membantu memahami kerangka kerja dan konsep yang relevan dalam mengidentifikasi pola atau tren yang mungkin ada[18].

2. Data Sekunder

Data sekunder atau data publik mengacu pada informasi yang diperoleh dari pihak atau peneliti lain. Pada penelitian ini dataset diperoleh dari website opendata.jabarprov.go.id. Data tersebut merupakan kumpulan data mengenai jumlah korban kekerasan berdasarkan tingkat pendidikan, jenis kelamin, dan kabupaten/kota di provinsi Jawa Barat.

3.2. Analisis Data

Algoritma *k-means clustering* diimplementasikan sebagai metode analisis data pada penelitian ini. Proses dilakukan dengan menentukan nilai *K* yang akan dibentuk, dan ditentukan secara acak. Dalam penelitian ini akan membentuk model *cluster* dari *K*=2 sampai *K*=10. Pembentukan model *cluster* dari rentang *K*=2 sampai *K*=10, dimaksudkan untuk melakukan uji coba dengan jumlah *cluster* yang berbeda-beda, serta diharapkan dapat memberikan fleksibilitas untuk mengeksplorasi berbagai kemungkinan jumlah *cluster*. Kemudian model *cluster* *K*=2 sampai *K*=10 akan di evaluasi untuk mengetahui jumlah *cluster* optimal yang terbentuk, dengan menggunakan metrik evaluaasi *Davies Bouldin Index (DBI)*. Setelah mengetahui jumlah *cluster* optimal, langkah selanjutnya adalah melakukan identifikasi pola dan karakteristik dalam data untuk memahami perbedaan dan kesamaan antara kelompok korban kekerasan di Jawa Barat.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Hasil analisis menggunakan algoritma *k-means clustering* dan pendekatan *Knowledge Discovery in Database(KDD)* pada penelitian ini diperoleh temuan:

1. Data

Data yang digunakan berasal dari Dinas Pemberdayaan Perempuan Perlindungan Anak dan Keluarga Berencana (DP3AKB), diunduh dari

diterbitkan antara tahun 2018 - 2022. Dataset ini mencakup informasi mengenai jumlah korban kekerasan berdasarkan tingkat pendidikan, jenis kelamin, dan Kabupaten/Kota di Provinsi Jawa Barat, terdiri dari 2.214 entri dan 10 kolom atribut. Selanjutnya, data dari tahun 2018 hingga

2022 digabungkan, menghasilkan dataset baru dengan 432 entri dan 10 atribut. Dengan penggabungan ini, total jumlah korban kekerasan mencapai 6.934 korban per tahun 2018-2022. Dataset hasil penggabungan dapat dilihat pada tabel 1.

Tabel 1. Isi dataset

Id	Kode Provinsi	Nama Provinsi	Kode Kabupaten/Kota	Nama Kabupaten/Kota	Jenis Kelamin	Kategori Pendidikan	Jumlah	Satuan	Tahun
1	32	Jawa Barat	3201	Kabupaten Bogor	Laki-Laki	NA	4	Korban	2018-2022
2	32	Jawa Barat	3201	Kabupaten Bogor	Laki-Laki	Tidak Sekolah	7	Korban	2018-2022
3	32	Jawa Barat	3201	Kabupaten Bogor	Laki-Laki	SD	3	Korban	2018-2022
4	32	Jawa Barat	3201	Kabupaten Bogor	Laki-Laki	SLTP	5	Korban	2018-2022
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
429	32	Jawa Barat	3279	Kota Banjar	Perempuan	SLTA	3	Korban	2018-2022
430	32	Jawa Barat	3279	Kota Banjar	Perempuan	Perguruan Tinggi	1	Korban	2018-2022
431	32	Jawa Barat	3279	Kota Banjar	Perempuan	TK	0	Korban	2018-2022
432	32	Jawa Barat	3279	Kota Banjar	Perempuan	PAUD	1	Korban	2018-2022

2. Selection

Tahap *selection* dilakukan untuk memilih dataset yang akan dianalisis dengan mengimpor data melalui operator *Read Excel* dalam *RapidMiner*. Operator *Read Excel* berperan membaca informasi data dari file tersimpan di repositori dan memuatnya ke dalam tab proses agar hasilnya dapat dilihat. Gambar operator dan hasil *Read Excel* dapat dilihat pada gambar 2.



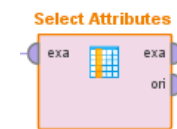
Gambar 2. Operator Read Excel

Tahap selanjutnya melibatkan penggunaan operator *Set Role*. Operator *Set Role* digunakan untuk menetapkan peran (role) setiap atribut dalam dataset. Pada tahap ini, atribut "id" ditetapkan sebagai target *role* untuk setiap atribut dalam dataset. Tampilan operator *Set Role* dapat dilihat pada gambar 3.



Gambar 3. Operator Set Role

Setelah menentukan atribut id sebagai *role*, tahap selanjutnya adalah pemilihan atribut menggunakan operator *Select Attributes*. Operator *Select Attributes* dapat dilihat pada gambar 4.



Gambar 4. Operator Select Atribut

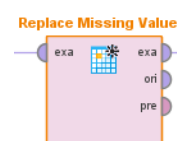
Pada operator *Select Attributes* proses seleksi dilakukan untuk menentukan atribut yang akan digunakan dalam proses analisis, hasil seleksi atribut dapat dilihat pada tabel 2.

Tabel 2. Hasil seleksi atribut

Atribut yang digunakan	Atribut yang tidak digunakan
Jenis Kelamin	Id
Jumlah	Kode Kabupaten/Kota
Kategori Pendidikan	Kode Provinsi
Nama Kabupaten/Kota	Nama Provinsi
-	Satuan
-	Tahun

3. Preprocessing

Pada tahap *Preprocessing*, diterapkan operator *Replace Missing Value* untuk mengatasi nilai kosong atau hilang pada atribut, dengan melakukan penghapusan data yang *missing*. Operator *Replace Missing Value* dan hasil proses *Replace Missing Value* dapat dilihat pada gambar 5 dan tabel 3.



Gambar 5. Operator Replace Missing Value

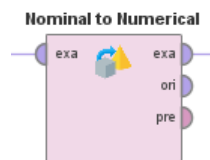
Tabel 3. Hasil pembersihan data

Nama Atribut	Missing
Jenis Kelamin	0
Jumlah	0
Kategori Pendidikan	0
Nama Kabupaten/Kota	0

Hasil menunjukan tidak ada nilai yang hilang dalam dataset. Oleh karena itu, tidak diperlukan tindakan penggantian atau imputasi data pada tahap ini.

4. Transformation

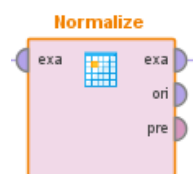
Tahap transformasi, operator *Nominal to Numerical* diterapkan dengan tipe pengkodean *unique integer* untuk mengubah atribut nominal menjadi numerik, untuk menyesuaikan tipe data yang digunakan oleh algoritma *k-means* pada proses pengelompokan di tahap *Data Mining*.

Gambar 6. Operator *Nominal to Numerical*

Tabel 4. Hasil transformasi data

Atribut	Tipe Data
Jumlah	Integer
Jenis Kelamin	Numerik
Kategori Pendidikan	Numerik
Nama Kabupaten/Kota	Numerik

Setelah melakukan transformasi data, langkah berikutnya adalah normalisasi pada atribut jumlah. Normalisasi dilakukan untuk memastikan skala atau rentang nilai dalam atribut jumlah seragam, mencegah dominasi nilai dengan rentang besar dalam proses analisis. Proses normalisasi menggunakan operator *Normalize*, yang dapat dilihat pada gambar 7.

Gambar 7. Operator *Normalize*

5. Data Mining

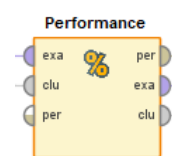
Data Mining digunakan untuk proses pencarian pola dan karakteristik dari dataset jumlah korban kekerasan berdasarkan kategori pendidikan, jenis kelamin, dan Kabupaten/Kota di provinsi Jawa Barat. Pengolahan data menggunakan tools *Rapidminer* dengan langkah sebagai berikut: Langkah pertama, terapkan operator *K-Means Clustering* untuk melakukan analisis dan pengelompokan data ke dalam beberapa cluster

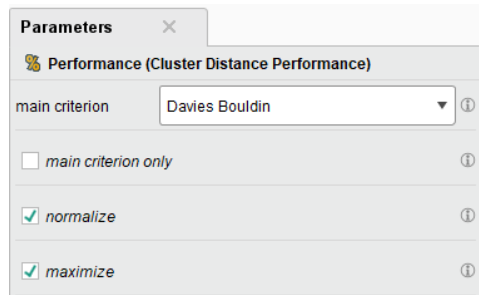
berdasarkan kesamaan karakteristik data. Proses pengujian dilakukan dengan membentuk model *cluster* dari K2 sampai K10. Pembentukan model *cluster* dengan rentang K=2 sampai K=10 dimaksudkan dapat memberikan fleksibilitas untuk mengeksplorasi berbagai kemungkinan jumlah *cluster*. Pengukuran menggunakan *Numerical Measures* serta *Euclidean Distance* untuk mengukur jarak antar titik data dan pusat *cluster*. Setiap pembentukan *cluster*, K dijalankan sebanyak 10 kali dengan inisialisasi acak yang berbeda, dan dilakukan iterasi maksimum 100 kali dalam satu percobaan dari algoritma *k-means* untuk memperbarui posisi *centroid*. Gambar operator *K-Means Clustering* dapat dilihat pada gambar 8.

Gambar 8. Operator *K-Means Clustering*

6. Evaluation

Setelah melakukan uji coba dengan membentuk K=2 sampai K=10, langkah selanjutnya adalah mengevaluasi kualitas *cluster* model yang terbentuk dengan menggunakan operator *Cluster Distance Performance*. Operator ini memanfaatkan model *centroid cluster* dan set data hasil pengelompokan sebagai *input*, kemudian mengevaluasi kinerja model berdasarkan *centroid cluster*. Evaluasi dilakukan dengan menetapkan parameter *Davies-Bouldin Index (DBI)* sebagai metrik evaluasi. *DBI* melibatkan perhitungan jarak *intra-cluster* dan *inter-cluster*. Dalam menghitung jarak *intra-cluster*, diukur jarak antara semua anggota dalam satu *cluster* untuk menilai sejauh mana tingkat kesamaan antar *cluster*. Hasil jarak *intra-cluster* memberikan gambaran tentang karakteristik serupa di dalam satu kelompok, dengan nilai rendah menunjukkan tingkat kesamaan yang tinggi. Sementara itu, dalam menghitung jarak *antar-cluster*, diukur jarak antara *centroid* (titik pusat) masing-masing kelompok untuk mengetahui sejauh mana satu kelompok berbeda dari kelompok lainnya. Jarak *antar-cluster* memberikan indikasi seberapa terpisah kelompok satu sama lain. Gambar operator dan parameter *Cluster Distance Performance* ditunjukkan pada gambar 9 dan gambar 10, serta hasil uji *performance DBI* pada tabel 5.

Gambar 9. Operator *Cluster Distance Performance*



Gambar 10. Parameter Cluster Distance Performance

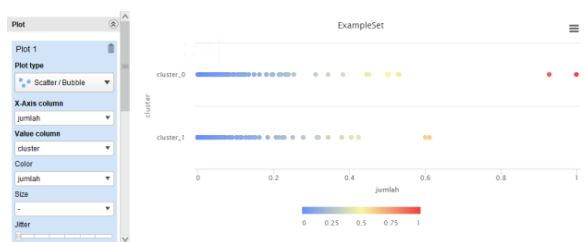
Tabel 5. Hasil uji performance DBI

Jumlah Cluster (K)	Performance Evaluation	Bavies Bouldin
2	DBI	0,157
3	DBI	0,183
4	DBI	0,213
5	DBI	0,250
6	DBI	0,251
7	DBI	0,245
8	DBI	0,254
9	DBI	0,224
10	DBI	0,222

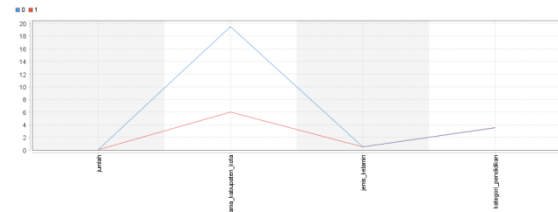
Pada tabel 5 ditunjukkan nilai DBI terkecil terdapat pada K=2 dengan nilai validitas sebesar 0,157.

Tabel 6. Hasil centroid K=2

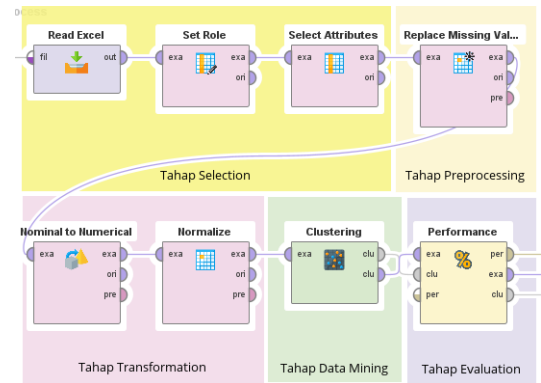
Atribut	Cluster 0	Cluster 1
Jenis Kelamin	0.5	0.5
Jumlah	0.0648938923395445 1	0.0508988294314381 35
Kategori Pendidikan	3.5	3.5528846153846154
Nama Kabupaten /Kota	19.5	6.0



Gambar 11. Visualisasi hasil clustering



Gambar 12. Grafik plot hasil clustering



Gambar 13. Alur tahapan KKD pada Rapidminer

4.2. Pembahasan

Berdasarkan tabel 5 hasil uji *performance*, dapat dilihat *cluster* dengan nilai DBI terkecil terdapat pada K=2 dengan nilai validitas sebesar 0.157. Analisis menunjukkan bahwa pembentukan 2 *cluster* memberikan hasil model yang optimal, dengan pola dan karakteristik:

- Cluster 0* mencakup Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Karawang, Kabupaten Pangandaran, Kabupaten Purwakarta, Kota Bandung, Kota Banjar, Kota Bekasi, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Depok, Kota Sukabumi, dan Kota Tasikmalaya, dimana korban dengan jenis kelamin laki-laki didominasi oleh tingkat pendidikan NA, tidak sekolah, SD, SLTA, dan TK dan korban dengan jenis kelamin perempuan didominasi oleh tingkat pendidikan NA, SD, SLTA, Perguruan Tinggi, TK, dan PAUD.
- Cluster 1* mencakup Kabupaten Bandung, Kabupaten Bogor, Kabupaten Ciamis, Kabupaten Cianjur, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Kuningan, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Sukabumi, Kabupaten Sumedang, dan Kabupaten Tasikmalaya, dimana korban dengan jenis kelamin laki-laki didominasi oleh tingkat pendidikan SLTP dan PAUD dan korban dengan jenis kelamin perempuan didominasi oleh tingkat pendidikan tidak sekolah dan SLTP. Pengelompokan *cluster 0* dan *cluster 1*, dapat dilihat pada tabel 7 dan tabel 8

Tabel 7. Hasil pengelompokan *cluster 0*

Nama Kabupaten/Kota	Jenis Kelamin	Kategori Pendidikan	Jumlah	Cluster
Kabupaten Purwakarta	Laki-laki	NA	19	0
Kabupaten Purwakarta	Laki-laki	Tidak Sekolah	0	0
Kabupaten Purwakarta	Laki-laki	SD	5	0
Kabupaten Purwakarta	Laki-laki	SLTP	2	0
Kabupaten Purwakarta	Laki-laki	SLTA	8	0
Kabupaten Purwakarta	Laki-laki	Perguruan Tinggi	2	0
Kabupaten Purwakarta	Laki-laki	TK	0	0
Kabupaten Purwakarta	Laki-laki	PAUD	0	0
.....				
.....				
Kota Banjar	Perempuan	NA	106	0
Kota Banjar	Perempuan	Tidak Sekolah	2	0
Kota Banjar	Perempuan	SD	4	0
Kota Banjar	Perempuan	SLTP	22	0
Kota Banjar	Perempuan	SLTA	3	0
Kota Banjar	Perempuan	Perguruan Tinggi	1	0
Kota Banjar	Perempuan	TK	0	0
Kota Banjar	Perempuan	PAUD	0	0

Tabel 8. Hasil pengelompokan *cluster 1*

Nama Kabupaten/Kota	Jenis Kelamin	Kategori Pendidikan	Jumlah	Cluster
Kabupaten Bogor	Laki-laki	NA	4	1
Kabupaten Bogor	Laki-laki	Tidak Sekolah	7	1
Kabupaten Bogor	Laki-laki	SD	3	1
Kabupaten Bogor	Laki-laki	SLTP	5	1
Kabupaten Bogor	Laki-laki	SLTA	3	1
Kabupaten Bogor	Laki-laki	Perguruan Tinggi	1	1
Kabupaten Bogor	Laki-laki	TK	0	1
Kabupaten Bogor	Laki-laki	PAUD	0	1
.....				
.....				
Kabupaten Subang	Perempuan	NA	24	1
Kabupaten Subang	Perempuan	Tidak Sekolah	0	1
Kabupaten Subang	Perempuan	SD	13	1
Kabupaten Subang	Perempuan	SLTP	25	1
Kabupaten Subang	Perempuan	SLTA	37	1
Kabupaten Subang	Perempuan	Perguruan Tinggi	1	1
Kabupaten Subang	Perempuan	TK	2	1
Kabupaten Subang	Perempuan	PAUD	0	1

5. KESIMPULAN DAN SARAN

Dari hasil penelitian diperoleh kesimpulan bahwa, algoritma *k-means* dapat diimplementasikan untuk mengelompokkan korban kekerasan di Provinsi Jawa Barat menggunakan tools *Rapidminer*. Pengujian dilakukan dengan membentuk model *cluster* K=2 sampai K=10 menggunakan tipe pengukuran *Numerical Measures* dan metrik jarak *Euclidean Distance*. Hasil pengujian menggunakan metrik evaluasi *Davies Bouldin Index (DBI)* dengan nilai validitas sebesar 0.157, diperoleh pada K=2, dengan karakteristik, *cluster 0* didominasi oleh korban dengan tingkat pendidikan NA, tidak sekolah, SD, SLTA, dan TK pada jenis kelamin laki-laki, dan tingkat pendidikan NA, SD, SLTA, Perguruan Tinggi, TK, dan PAUD pada korban dengan jenis kelamin perempuan. *Cluster 1* didominasi oleh korban dengan tingkat pendidikan SLTP dan PAUD

pada jenis kelamin laki-laki dan tingkat pendidikan tidak sekolah dan SLTP pada jenis kelamin perempuan. Adapun saran untuk penelitian selanjutnya yaitu diharapkan menambahkan data korban serta gunakan tools yang berbeda.

DAFTAR PUSTAKA

- [1] Noviya Adawiyah, Nina Sulistiyowat, and ohamad Jajuli, "Klasterisasi Kasus Kekerasan Terhadap Anak dan Perempuan Berdasarkan Algoritma *K-means* ," *Gener. J.*, vol. 5, no. 2, pp. 69–80, 2021.
- [2] R. Rahma and R. Mufidah, "Pengelompokan Daerah Rawan Kekerasan Terhadap Perempuan Dan Anak Di Jawa Barat Menggunakan Algoritma *K-means* ," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 3, pp. 850–857, 2022, doi: 10.29100/jupi.v7i3.3144.

- [3] M. Nandang, R. Resnawaty, and A. Wahyudi, "Pelayanan Sosial Bagi Anak Korban Kekerasan (Social Service for Child Abuse)," *Pros. Penelit. dan Pengabd. Kpd. Masy.*, vol. 6, no. 2, p. 123, 2019, doi: 10.24198/jppm.v6i2.19785.
- [4] Kemenpppa, "Data Kekerasan Nasional," *SIMFONI PPA (Sistem Informasi Online Perlindungan Perempuan dan Anak)*, 2023. <https://kekerasan.kemenpppa.go.id/ringkasan> (accessed Nov. 21, 2023).
- [5] Q. A'yuni, A. Nazir, L. Handayani, and I. Afrianty, "Penerapan Algoritma *K-means clustering* untuk Mengetahui Pola Penerima Beasiswa Bank Indonesia (BI)," *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 530–539, 2023, doi: 10.47065/josyc.v4i3.3343.
- [6] M. Herviany, S. Putri Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma *K-means* dan *K-Medoids* untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.
- [7] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.
- [8] D. A. T. Devanta, "Optimization of the *K-means clustering* Algorithm Using Davies Bouldin Index in Iris Data Classification," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 545–552, 2023, doi: 10.30865/klik.v4i1.964.
- [9] T. N. P. Halimatu Sa'diah, Ultach Enri, "PENERAPAN ALGORITME *K-MEANS* DALAM SEGMENTASI DAERAH RAWAN KEKERASAN ANAK DI JAWA BARAT," vol. 7, no. 2, pp. 1351–1357, 2023.
- [10] D. K. Sitinjak, B. A. Pangestu, and B. N. Sari, "Clustering Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma *K-means*," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 47–54, 2022, doi: 10.30871/jaic.v6i1.3855.
- [11] A. Wibowo, Moh Makruf, Inge Virdyna, and Farah Chikita Venna, "Penentuan Klaster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 565–575, 2021, doi: 10.29207/resti.v5i3.3041.
- [12] G. Triyandana, L. A. Putri, and Y. Umaidah, "Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode *K-means*," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 40–46, 2022, doi: 10.30871/jaic.v6i1.3824.
- [13] F. P. Ferdy Pangestu, N. Y. Nur Yasin, R. C. Ronald Chistover Hasugian, and Y. Yunita, "Penerapan Algoritma *K-means* Untuk Mengklasifikasi Data Obat," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 53–62, 2023, doi: 10.32736/sisfokom.v12i1.1461.
- [14] N. Buslim and R. P. Iswara, "Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis di Media Sosial sebagai media promosi Online Bagi Masyarakat," *J. Tek. Inform.*, vol. 12, no. 1, pp. 79–96, 2019, doi: 10.15408/jti.v12i1.11342.
- [15] T. Wahyudi and T. Silfia, "Implementation of Data Mining Using *K-means clustering* Method To Determine Sales Strategy in S&R Baby Store," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 1, pp. 93–103, 2022, doi: 10.37385/jaets.v4i1.913.
- [16] W. Wahyu Pribadi, A. Yunus, and A. S. Wiguna, "Perbandingan Metode *K-means* Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 493–500, 2022, doi: 10.36040/jati.v6i2.4808.
- [17] F. Achmad, O. Nurdian, and Y. Arie Wijaya, "Analisa Pola Transaksi Pembelian Konsumen Pada Toko Ritel Kesehatan Menggunakan Algoritma Fp-Growth," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 168–175, 2023, doi: 10.36040/jati.v7i1.6210.
- [18] M. A. Sundari, R. Pane, and R. Rohani, "Data Mining Clustering Korban Kejahatan Pelecehan Seksual dengan Kekerasan Berdasarkan Provinsi Menggunakan Metode AHC," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3499.