

PENERAPAN ALGORITME K-MEANS DALAM SEGMENTASI DAERAH RAWAN KEKERASAN ANAK DI JAWA BARAT

Halimatu Sa'diah, Ultach Enri, Tesa Nur Padilah

Program Studi Informatika S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
1910631170082@student.unsika.ac.id

ABSTRAK

Kekerasan pada anak merupakan segala bentuk perbuatan atau tindakan yang menimbulkan penderitaan baik secara fisik atau emosional, seksual, eksploitasi, penelantaran yang berdampak kerugian pada keberlangsungan hidup anak. Tingkat kekerasan anak di dunia mencapai 1 miliar anak dalam kurun waktu 1 tahun. Di Indonesia sendiri terjadi peningkatan jumlah korban setiap tahunnya. Hal ini tidak terlepas dari jumlah kasus yang terjadi di setiap kabupaten/kota pada setiap provinsi. Jawa Barat merupakan salah satu dari tiga provinsi dengan korban kekerasan terbanyak di Indonesia. Dalam membantu upaya penyelesaian kasus yang lebih terstruktur dibutuhkan pengendalian dan penekanan kasus kekerasan terhadap anak menggunakan segmentasi wilayah agar dapat membantu pemerintah dalam menentukan daerah prioritas penanganan. Penelitian ini memanfaatkan data dalam rentang tahun 2017-2021. Algoritme yang diterapkan adalah algoritme *K-Means* menggunakan metodologi KDD yaitu *data selection*, *data preprocessing*, *transformation*, *data mining*, dan *interpretation/evaluation*. Terbentuk 3 *cluster* dengan 4 daerah rawan, 1 sangat rawan, dan 22 tidak rawan. *Cluster* menghasilkan 0.36 dengan *silhouette coefficient* 0.76 (*strong structure*). Interpretasi divisualisasikan menggunakan QGIS.

Kata kunci : *K-Means, Clustering, Kekerasan Anak.*

1. PENDAHULUAN

Kekerasan pada anak merupakan segala bentuk perbuatan atau tindakan yang menimbulkan penderitaan baik secara fisik atau emosional, seksual, eksploitasi, penelantaran yang berdampak kerugian pada keberlangsungan hidup anak [1]. Tingkat kekerasan pada anak di dunia secara global berdasarkan sumber yang diambil dari *website* resmi *World Health Organization* (WHO) mencapai 1 miliar anak mengalami kekerasan dalam kurun waktu 1 tahun terakhir [2]. Sedangkan menurut data dari *United Nations International Children's Emergency Fund* (UNICEF) pada 129 negara yang didukung UNICEF pada tahun 2021 sebanyak 4,4 juta anak pernah mengalami kekerasan dan hal ini merupakan peningkatan sebanyak 80% dari tahun 2017 [3]. Berdasarkan data di atas telah terjadi peningkatan kekerasan anak pada skala global yang terjadi pada berbagai negara di dunia.

Pada negara Indonesia terjadi peningkatan serupa terhitung menurut Kementerian Pemberdayaan Perempuan dan Anak (Kemen PPPA) terjadi peningkatan kekerasan anak yaitu Jumlah kasus kekerasan terhadap anak yang dilaporkan meningkat dari 11.057 kasus pada tahun 2019, kemudian 11.278 kasus pada tahun 2020 menjadi 14.517 kasus pada tahun 2021. Sedangkan untuk jumlah korban kekerasan terhadap anak juga meningkat dari 12.285 pada tahun 2019 menjadi 12.425 pada tahun 2020 dan sejumlah 15.972 pada 2021 [4].

Provinsi Jawa Barat merupakan salah satu dari tiga provinsi dengan korban kekerasan terbanyak di Indonesia dengan jumlah 1766 kasus di Jawa Barat,

2162 kasus di Jawa Tengah, dan 2144 kasus di Jawa Timur [5].

Berdasarkan data di atas kekerasan terhadap anak yang kian meningkat mencerminkan bahwa permasalahan ini belum ditangani dengan baik. Permasalahan tersebut tidak bisa dibiarkan dan harus diselesaikan secara lebih terstruktur karena kekerasan terhadap anak dapat menimbulkan berbagai dampak. Dalam membantu upaya penyelesaian kasus yang lebih terstruktur dibutuhkan pengendalian dan penekanan kasus kekerasan terhadap anak menggunakan segmentasi wilayah agar dapat membantu pemerintah dalam menentukan daerah prioritas penanganan. Klasterisasi yang dilakukan didasarkan pada tingkat kekerasan anak yang berbeda-beda pada setiap daerah agar pemerintah mampu memberikan tindakan yang tepat dan sesuai dengan kondisi di tiap-tiap wilayah serta penanganan yang dilakukan pun akan lebih terfokus.

Berdasarkan permasalahan tersebut maka pada penelitian ini akan dilakukan segmentasi daerah rawan kekerasan terhadap anak pada kabupaten/kota di Jawa Barat berdasarkan bentuk kekerasan dan jumlah kasus. Dengan melakukan analisis pada hasil segmentasi diharapkan pemerintah Jawa Barat dan lembaga yang berfokus pada kesejahteraan anak dapat menentukan langkah yang tepat dalam menekan kasus kekerasan pada anak dalam lingkup kabupaten/kota di Jawa Barat. Algoritma yang akan digunakan yaitu menggunakan algoritma *k-means* yang akan dilakukan pengoptimalan menggunakan metode *elbow* dan *silhouette coefficient* serta dilakukan visualisasi dengan menggunakan *tools Quantum Geographic Information System (QGIS)*.

2. TINJAUAN PUSTAKA

2.1. Kekerasan

Kekerasan pada arti sempit mengacu pada tindakan penyerangan, perusakan, penghancuran diri sendiri secara fisik atau pada sesuatu yang berpotensi milik orang lain. Sedangkan Kekerasan secara luas mengacu pada tindakan fisik atau psikologis yang dilakukan seorang individu atau gerombolan individu secara sengaja atau tidak sengaja, eksklusif atau tidak eksklusif [6].

2.2. Kekerasan Anak

Kekerasan pada anak merupakan segala bentuk perbuatan atau tindakan yang menimbulkan penderitaan baik secara fisik atau emosional, seksual, eksploitasi, penelantaran yang berdampak kerugian pada keberlangsungan hidup anak [1].

2.3. Data Mining

Data mining merupakan proses penambangan informasi dari sekumpulan data dalam *database* yang besar dan kompleks dengan cara mengekstraksi data dan yang tersedia sehingga menjadi pengetahuan dan berguna dalam pengambilan keputusan [7]. *Data mining* digunakan dalam berbagai bidang, seperti bisnis, ilmu pengetahuan, teknologi informasi, dan sebagainya untuk menemukan pola dalam data [8].

2.4. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah kumpulan tahapan dalam proses identifikasi data menjadi kumpulan informasi yang mengandung pola. Informasi ini terdapat dalam basis data berukuran besar yang sebelumnya belum diketahui manfaat potensial dari data tersebut [9]. Tahapan dari KDD diantaranya yaitu *data selection*, *data preprocessing*, *transformation*, *data mining*, dan *evaluation* [10].

2.5. Clustering

Clustering adalah pengelompokan objek berupa data dengan karakteristik atau sifat yang mirip ke dalam sebuah *cluster* dan objek atau data dengan tingkat kemiripan yang rendah terhadap *cluster* yang lain dengan melakukan pemaksimalan kesamaan antar objek dalam sebuah klaster dan meminimalkan kesamaan antar *cluster* [11]. *Clustering* merupakan metode atau teknik menambang data (*data mining*) yang digunakan untuk menganalisa data [12].

2.6. K-Means

K-Means merupakan sebuah metode dalam penambangan data yang bermanfaat untuk mengelompokkan data yang tidak memiliki tingkatan atau hanya bersifat partisi. Metode ini dapat membagi objek menjadi kelompok-kelompok, yang dapat membagi data menjadi satu kelompok yang mempunyai kemiripan dan memisahkannya dengan kelompok lain. Metode ini didasarkan terhadap jarak antar data di dalam *cluster* yang terbentuk. Semakin kecil jarak maka semakin mirip antara data satu

dengan data lainnya. Algoritma K-Means memiliki beberapa kelebihan yaitu mampu mengagregasi data dalam objek yang besar dengan proses yang singkat dan efisien [13].

2.7. Metode Elbow

Metode *elbow* adalah metode untuk mengoptimalkan pemilihan titik *cluster* atau *centroid* dalam proses penambangan data. Optimalisasi dilakukan dengan memilih nilai *k* yang kecil dan yang memiliki nilai *witness* rendah. Pada metode *elbow cluster* optimal akan mengalami penurunan nilai SSE yang besar setelah itu titik-titik berikutnya penurunan akan cenderung stabil [14].

2.8. Silhouette Coefficient

Jarak antar data pada *silhouette coefficient* dihitung menggunakan rumus *euclidian distance* [15]. Kriteria pengukuran pengelompokan berdasarkan score *silhouette coefficient* disajikan pada Tabel 2.1.

Tabel 1. Kriteria Pengukuran Silhouette Coefficient

Score	Kriteria
0.71-1.00	Struktur Kuat
0.51-0.70	Struktur Baik
0.26-0.50	Struktur Lemah
≤ 0.25	Struktur Buruk

2.9. Sum of Square Error

Sum of Square Error (SSE) adalah rumus yang diterapkan untuk mengukur perbedaan antara data yang didapat dengan model perkiraan yang telah dilakukan sebelumnya [16].

2.10. Davies Bouldin Index

Davies Bouldin Index (DBI) adalah salah satu metode validasi hasil *clustering*. *Cluster* dianggap optimal dan berkualitas jika memiliki indeks DBI yang semakin rendah [17].

2.11. Rstudio

Rstudio merupakan *Integrated Development Environment (IDE)* untuk tampilan antarmuka dari *R Programming*. *RStudio* adalah perangkat lunak yang *open source* yang dapat dijalankan pada *desktop* atau *browser* yang terhubung dengan *server RStudio* [18]. Dalam pemilihan jumlah *cluster* yang optimal dapat menggunakan *tools RStudio* [19].

2.12. Quantum Geographic Information System

Sistem Informasi Geografis (SIG) adalah sebuah sistem yang berguna untuk proses pengumpulan, pemeriksaan, dan analisis informasi yang berkaitan dengan kondisi geografis permukaan bumi. *Quantum Geographic Information System (QGIS)* adalah *software SIG* secara terbuka dan tidak memerlukan biaya yang biasanya ditujukan untuk pengolahan data geospasial [20].

3. METODE PENELITIAN

Metodologi penelitian ini yaitu menggunakan metode *Knowledge Discovery in Database (KDD)*. KDD merupakan kumpulan tahapan dalam proses identifikasi data menjadi kumpulan informasi yang mengandung pola. Informasi ini terdapat dalam basis data berukuran besar yang sebelumnya belum diketahui manfaat potensial dari data tersebut. Proses yang dilakukan yaitu melalui tahapan *data selection*, *data preprocessing*, *transformation*, *data mining*, dan *interpretation/evaluation*.



Gambar 1. Rancangan Penelitian

4. HASIL DAN PEMBAHASAN

Pada tahap *data selection* dilakukan proses pengumpulan data mentah atau *dataset* awal. Pada tahap ini dilakukan pengumpulan data, *dataset* yang digunakan merupakan data publik yang diambil dari website.opendata.jabarpov.go.id yang dihimpun oleh

Dinas Pemberdayaan Perempuan Perlindungan Anak dan Keluarga Berencana (DP3AKB). Data yang diambil merupakan data jumlah korban kekerasan pada anak perempuan dan anak laki-laki berdasarkan bentuk kekerasan di Jawa Barat pada rentang waktu 2017-2021 dengan masing-masing memiliki 945 baris dan 9 kolom dengan keterangan atribut awal yaitu id, nama_provinsi, kode_provinsi, kode_kabupaten_kota, nama_kabupaten, bentuk_kekerasan, jumlah_korban, satuan, dan tahun.

Pada tahap selanjutnya yaitu data *preprocessing* atau pemrosesan. Langkah pertama adalah dengan melakukan penghapusan pada data yang tidak diperlukan. Data dianggap tidak diperlukan karena tidak berhubungan dengan penelitian atau atribut tersebut tidak berpengaruh terhadap proses analisis data. Atribut yang dihapus yaitu atribut id, kode_provinsi, nama_provinsi, kode_kabupaten_kota, satuan, dan atribut tahun. Selain itu dilakukan pula perubahan pada struktur tabel dengan mengubah kolom yang memiliki atribut bentuk kekerasan diubah menjadi 7 kolom dengan label baru yang sebelumnya merupakan isi dari kolom bentuk kekerasan serta dilakukan juga penambahan 1 kolom dengan label total korban. Setelah data diintegrasikan, Langkah selanjutnya yaitu melakukan pengecekan *missing value*.

```

> is.na(tabelfix)
      EKSPLOITASI FISIK LAZINNYA PENELANTARAN PSIKIS SEKSUAL TRAFFICKING TOTAL_KORBAN
KABUPATEN BOGOR          FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN SUKABUMI        FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN CIANJUR         FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN BANDUNG         FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN GARUT           FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN TASIKMALAYA     FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN CIPARIS         FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN KINERJAS        FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN CIREBON         FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN MAJALENGKA      FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN SURENDRANG      FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN INDRAMAYU       FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN SUBANG          FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN PURBAKARTA      FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN KUNING          FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN BEKASI          FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN BANDUNG BARAT   FALSE FALSE          FALSE FALSE          FALSE          FALSE
KABUPATEN PANJANDARAN     FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA BOGOR                FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA SUKABUMI             FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA BANDUNG              FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA CIREBON              FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA BEKASI               FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA DEPOK                FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA CIMAHI               FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA TASIKMALAYA          FALSE FALSE          FALSE FALSE          FALSE          FALSE
KOTA BANJAR               FALSE FALSE          FALSE FALSE          FALSE          FALSE
  
```

Gambar 2. Pengecekan *missing value*

Pada Gambar 2. tidak terdapat nilai *TRUE* yang menunjukkan bahwa tidak terdapat data yang memiliki nilai kosong. Maka dari itu dapat dilanjutkan ke pengecekan selanjutnya yaitu pengecekan redundansi data atau data yang terindikasi memiliki duplikat.

Setelah pengecekan data duplikat telah selesai dilakukan. Langkah selanjutnya yaitu tahap transformasi data. Pada tahap ini dilakukan perubahan data kompleks menjadi lebih mudah diolah menggunakan fungsi penskalaan.

```

#proses transformasi |
tabelfix=scale(tabelawal[,2:9])
view(tabelfix)
  
```

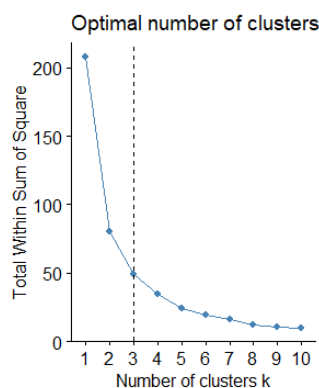
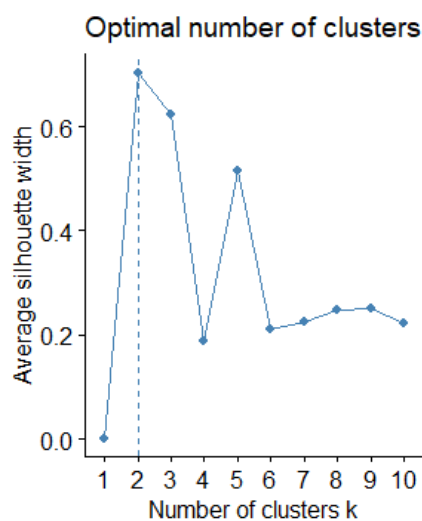
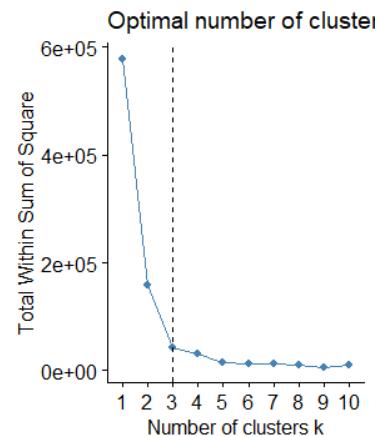
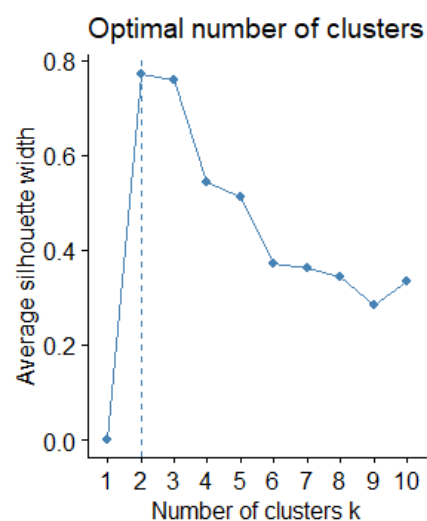
Gambar 3. Proses penskalaan

Standarisasi data dilakukan jika data memiliki banyak variabel dengan rentang nilai yang cukup jauh. Tujuan dari standarisasi data yaitu agar data dengan

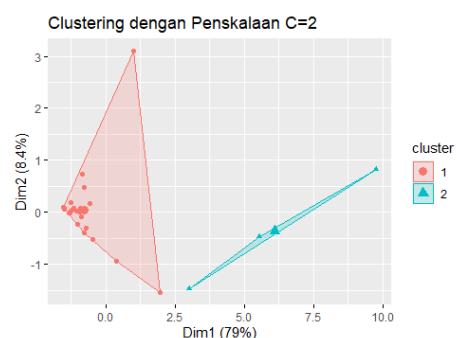
variabel berbeda memiliki rentang nilai yang sama. Untuk menghindari ketimpangan yang menyebabkan model clustering menjadi tidak optimal maka dilakukan klasterisasi data dengan penskalaan dan tanpa penskalaan.

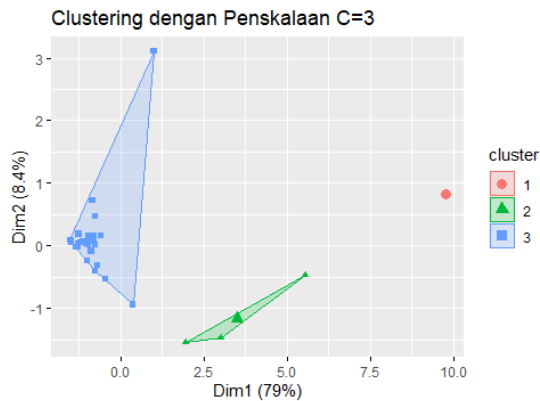
Setelah data siap diolah, tahapan selanjutnya yaitu proses data mining untuk melihat pola yang terdapat pada data. Teknik yang diterapkan pada proses ini adalah teknik *clustering* yaitu melakukan pengelompokan atau segmentasi data berdasarkan tingkat kemiripan dalam *cluster*. Algoritma yang digunakan yaitu algoritma *K-Means* dengan optimalisasi penentuan jumlah *cluster* menggunakan metode *elbow* yang didasarkan pada hasil perhitungan *Sum of Square* dan juga grafik *silhouette* digunakan sebagai pertimbangan penentuan jumlah *cluster*.

Metode *elbow* merujuk pada nilai k yang kecil serta memiliki nilai *witness* rendah. Titik yang menunjukkan *cluster* optimal dalam metode ini akan membentuk siku dalam sebuah grafik. Pembentukan titik optimal didasarkan pada perbandingan selisih penurunan pada nilai *sum of squared (SSE)* yang stabil dan tidak mengalami banyak perubahan pada *cluster* atau titik berikutnya.

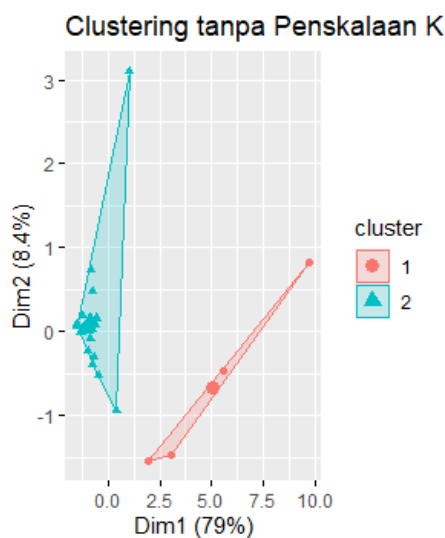
Gambar 4. Grafik *elbow* dengan penskalaanGambar 5. Grafik *silhouette* dengan penskalaanGambar 6. Grafik *elbow* tanpa penskalaanGambar 7. Grafik *silhouette* tanpa penskalaan

Dalam penentuan jumlah *cluster* berdasarkan metode *elbow cluster* optimal terdapat pada $C=3$ sedangkan pada grafik *silhouette coefficient cluster* optimal berada pada titik $C=2$ dan titik optimal selanjutnya yaitu $C=3$. Untuk itu, dilakukan proses clustering dengan jumlah cluster $C=2$ dan $C=3$ baik pada data dengan penskalaan dan data tanpa penskalaan.

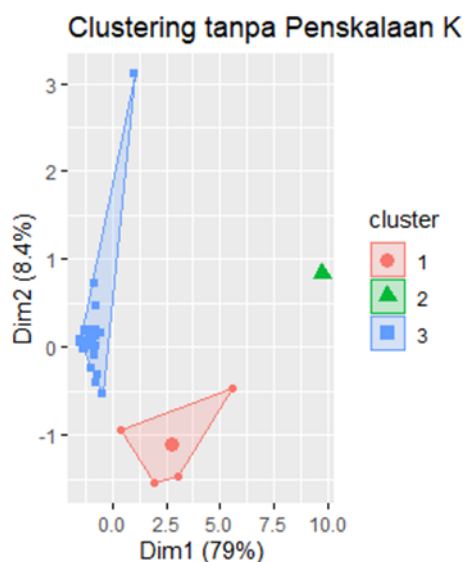
Gambar 8. *Clustering* $C=2$ dengan penskalaan



Gambar 9. Clustering C=3 dengan penskalaan



Gambar 10. Clustering C=2 tanpa penskalaan



Gambar 11. Clustering C=3 tanpa penskalaan

Setelah melalui tahapan data mining dan telah dilakukan segmentasi cluster. Langkah selanjutnya yaitu melakukan tahap evaluasi untuk mengukur hasil atau kualitas dari *cluster* dan model yang dibentuk.

Karena pada penentuan jumlah *cluster* terdapat perbedaan antara hasil dari grafik *elbow* di mana jumlah *cluster* optimal terletak pada titik C=3 dan pada grafik *silhouette coefficient* di mana jumlah *cluster* optimal pertama terletak pada titik C=2 dan *cluster* optimal kedua terletak pada titik C=3. Maka dilakukan evaluasi menggunakan nilai *Sum of Square (SSE)* yang diperkuat dengan *Davies Bouldin Index*.

Tabel 2. Hasil indeks *clustering* dengan penskalaan

Metode	Jumlah Cluster	SSE/ Tot. Witthniss	Betweeness
<i>Silhouette</i>	2	80.3871	127.6129.
<i>Elbow</i>	3	49.20053	158.7995

Tabel 3. Hasil indeks *clustering* tanpa penskalaan

Metode	Jumlah Cluster	SSE/ Tot. Witthniss	Betweeness
<i>Silhouette</i>	2	158392.2.	419941.5
<i>Elbow</i>	3	43582.41	534751.3

Indeks SSE atau dalam Rstudio ditunjukkan menggunakan fungsi *Tot. Witthniss*. *Cluster* dinyatakan optimal atau baik jika memiliki rasio nilai SSE yang rendah atau kecil. Sedangkan untuk indeks *betweeness*, sebuah *cluster* dinyatakan optimal jika memiliki rasio yang lebih tinggi atau semakin besar nilainya. Berdasarkan Tabel 2. dan Tabel 3. jumlah *cluster* yang memiliki nilai SSE lebih rendah nilai dan *betweeness* yang besar terletak pada titik C=3. Hal tersebut berlaku pada kedua skenario baik pada data dengan skala dan tanpa skala memiliki titik *cluster* optimal yang sama. Untuk memperkuat hasil analisis dilakukan evaluasi juga menggunakan *Davies Bouldin Index (DBI)*. Tabel perbandingan DBI dapat dilihat pada Tabel 4.

Tabel 4. Perbandingan DBI

Skenario	Metode	Jumlah Cluster	DBI
Data melalui penskalaan	<i>Silhouette</i>	2	0.6821
	<i>Elbow</i>	3	0.6072
Data tanpa penskalaan	<i>Silhouette</i>	2	0.6504
	<i>Elbow</i>	3	0.3695

Pada *Davies Bouldin Index (DBI)*, pengelompokan data dinyatakan optimal atau semakin baik jika memiliki *index* yang lebih rendah atau mendekati angka 0. Berdasarkan Tabel 3, kedua skenario memiliki *index* yang lebih rendah pada C=3. Maka titik optimal *cluster* baik berdasarkan nilai SSE dan nilai DBI berada di C=3. Setelah mengetahui bahwa pada kedua skenario titik optimal berada pada C=3. Maka langkah selanjutnya adalah membandingkan kedua model dengan menggunakan indeks *silhouette coefficient* untuk mengetahui kualitas dan kekuatan dari struktur model *clustering*.

Hasil menunjukan bahwa *indeks silhouette coefficient* pada data yang telah melalui tahap penskalaan adalah 0.62 yang berarti model segmentasi

tersebut memiliki struktur yang baik atau sedang karena berada pada rentang 0.51-0.70. Sedangkan pada data yang tanpa melalui tahap penskalaan memperoleh indeks *silhouette coefficient* sebesar 0.76 yang berarti model segmentasi tersebut memiliki struktur yang kuat karena berada pada rentang 0.71-1.00. Sehingga untuk penyajian atau interpretasi dari penelitian ini akan menggunakan data yang tidak menggunakan fungsi *scale* dengan jumlah *cluster* $C=3$. Untuk tabel hasil *clustering* dapat dilihat pada Tabel 5 dan visualisasi peta segmentasi daerah rawan kekerasan anak dapat dilihat pada Gambar 12.

Tabel 5. Hasil *clustering*

C1	C2	C3
Kabupaten Sukabumi	Kota Bandung	Kabupaten Bogor
Kabupaten Bandung		Kabupaten Cianjur
Kabupaten Tasikmalaya		Kabupaten Garut
Kabupaten Cirebon		Kabupaten Ciamis
		Kabupaten Kuningan
		Kabupaten Majalengka
		Kabupaten Sumedang
		Kabupaten Indramayu
		Kabupaten Subang
		Kabupaten Purwakarta
		Kabupaten Karawang
		Kabupaten Bekasi
		Kabupaten Bandung Barat
		Kabupaten Pangandaran
		Kota Bogor
		Kota Sukabumi
		Kota Cirebon
		Kota Bekasi
		Kota Depok
		Kota Cimahi
		Kota Tasikmalaya
		Kota Banjar



Gambar 12. Peta segmentasi daerah rawan kekerasan anak di Jawa Barat.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat diperoleh kesimpulan bahwa cluster terbaik terletak pada $C=3$ dengan data tanpa menggunakan fungsi *scale*. Hasil *clustering* mempunyai indeks SSE 43582.41 dan indeks *between*s 534751.3. Hal ini juga diperkuat dengan nilai DBI dengan *score* 0.3695. *Score silhouette* sebesar 0.76 menunjukkan bahwa model *cluster* pada data tanpa fungsi *scale* memiliki struktur yang kuat (*strong structure*), adapun saran untuk pengembangan penelitian yaitu diharapkan dapat menggunakan algoritma dan tools yang berbeda dengan variable yang lebih banyak.

DAFTAR PUSTAKA

- [1] T. S. Margareta and M. P. S. J. Jaya, "Kekerasan pada Anak (Tri Sella, Melinda P.S Jaya)," *Wahana Didakt.*, vol. 18, no. 2, pp. 171–180, 2020.
- [2] "Violence Against Children," *World Health Organization*, 2022. <https://www.who.int/news-room/fact-sheets/detail/violence-against-children> (accessed Jan. 04, 2022).
- [3] UNICEF, "Every child is protected from violence and exploitation," *Glob. Annu. Results Rep.* 2021, pp. 1–137, 2021.
- [4] A. Ramadhan, "Laporan Kasus Kekerasan terhadap Anak dan Perempuan Meningkat 3 Tahun Terakhir," *Kompas*, 2022. <https://nasional.kompas.com/read/2022/01/20/12435801/laporan-kasus-kekerasan-terhadap-anak-dan-perempuan-meningkat-3-tahun> (accessed Jan. 04, 2022).
- [5] S. D. Caesaria, "3 Provinsi dengan Angka Kekerasan Anak Tertinggi di Indonesia," *Kompas*, 2022. <https://www.kompas.com/edu/read/2022/05/31/135831971/3-provinsi-dengan-angka-kekerasan-anak-tertinggi-di-indonesia?page=all> (accessed Jan. 04, 2022).
- [6] E. M. Setiadi, *Pengantar ringkas sosiologi : pemahaman fakta dan gejala permasalahan sosial (teori, aplikasi, dan pemecahannya)*, 1st ed. Jakarta: Kencana, 2020.
- [7] J. Sutoro, *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo, 2019.
- [8] M. R. Nahjan, N. Heryana, and A. Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 1–4, 2023.
- [9] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [10] A. Fahllevi, "Proses Data Mining KDD," *Binus*

- University, 2022.
<https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/> (accessed Jan. 09, 2023).
- [11] I. Nuryani and D. Darwis, "Analisis Clustering Pada Pengguna Brand Hp Menggunakan Metode K-Means," *Pros. Semin. Nas. Ilmu Komput.*, vol. 1, no. 1, p. 2021, 2021.
- [12] B. Nurseptia, A. Voutama, N. Haeryana, and J. HSRonggo Waluyo, "Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru," *J. Teknol. Informasi*, vol. 6, no. 2, pp. 181–186, 2022.
- [13] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [14] A. Ardiansyah and F. Nggeboe, "Kajian Yuridis Penelantaran Anak Oleh Orang Tua Menurut Perspektif Hukum Indonesia," *Leg. J. Huk.*, vol. 10, no. 1, p. 146, 2019, doi: 10.33087/legalitas.v10i1.160.
- [15] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [16] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-means Menggunakan Metode Elbow pada Data Pengguna Narkoba dengan Pemrograman Python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [17] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [18] M. R. Faisal and D. T. Nugraha, *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*, Cetakan 1. Banjarbaru: Scriptaka Cendekia, 2019.
- [19] V. Ramadhan and A. Voutama, "Clustering Menggunakan Algoritma K-Means Pada Penyakit ISPA di Puskesmas Kabupaten Karawang," *J. Pendidik. dan Konseling*, vol. 4, pp. 462–473, 2022.
- [20] S. Bahri, D. M. Midyanti, and R. Hidayati, "Pemanfaatan QGIS Untuk Pemetaan Fasilitas Layanan Masyarakat Di Kota Pontianak," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 1, p. 70, 2020, doi: 10.24114/cess.v5i1.15666.