

**ANALISIS KLASSTER KASUS KEKERASAN ANAK DI SUMATERA SELATAN
BERBASIS PEMROGRAMAN FUNGSI R KELAS S3
KOMPUTASI STATISTIK**

Kelompok 9

Nama Anggota Kelompok:

- | | |
|------------------------------------|-----------|
| 1. Eigi Artamevia | 123450011 |
| 2. Muhammad Ridwan | 123450091 |
| 3. Nadia Faraj Alyafaatin Simbolon | 123450092 |
| 4. Arienta Khusnul Ananda | 123450097 |



**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
2025**

Abstrak

Penelitian ini bertujuan untuk menganalisis pola kasus kekerasan terhadap anak di Provinsi Sumatera Selatan pada tahun 2023 serta mengelompokkan kabupaten/kota berdasarkan karakteristik jenis kekerasan yang dominan. Data yang digunakan merupakan data sekunder dari Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KemenPPPA) yang mencakup tujuh kategori kekerasan, yaitu fisik, psikis, seksual, eksploitasi, TPPO, penelantaran, dan kategori lainnya. Metode K-Means *Clustering* diterapkan menggunakan pemrograman R dengan pendekatan Class S3 untuk menghasilkan proses analisis yang terstruktur. Tahapan penelitian meliputi pembersihan dan normalisasi data, *cross validation* untuk penentuan kestabilan jumlah klaster, pengelompokan menggunakan K-Means, serta visualisasi hasil melalui bar chart, heatmap, PCA, dan peta spasial. Hasil analisis menunjukkan terbentuknya tiga klaster utama: klaster kasus rendah, klaster kasus sedang, dan klaster kasus tinggi, dengan klaster ketiga didominasi oleh tingginya kasus kekerasan seksual dan psikis. Pemetaan spasial menunjukkan distribusi kerawanan yang tidak merata antarwilayah dan menegaskan perlunya pendekatan kebijakan yang berbeda pada setiap klaster. Secara keseluruhan, penelitian ini memberikan gambaran menyeluruh mengenai pola kekerasan anak di Sumatera Selatan dan menunjukkan bahwa metode K-Means berbasis pemrograman fungsi R dapat menjadi alat analitis yang efektif dalam mendukung perumusan strategi pencegahan serta penanganan kekerasan anak secara lebih tepat sasaran.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Fenomena kekerasan terhadap anak merupakan masalah sosial, hukum, dan kemanusiaan yang kompleks serta membutuhkan perhatian serius dari berbagai pihak. Dampak yang ditimbulkan oleh kekerasan pada anak sangat merugikan, tidak hanya dari aspek fisik, tetapi juga psikis, dan dapat mengganggu tumbuh kembang serta masa depan korban[1]. Berdasarkan data yang ada, kasus kekerasan terhadap anak di Provinsi Sumatera Selatan menunjukkan variasi yang signifikan antar wilayah. Jenis kekerasan yang dialami, seperti kekerasan fisik, psikis, dan seksual, memiliki distribusi yang berbeda-beda di setiap kabupaten/kota. Perbedaan pola kasus ini mengindikasikan bahwa faktor pemicu dan tingkat kerawanan di setiap daerah tidaklah sama, sehingga membutuhkan strategi penanganan yang terpersonalisasi.

Untuk memahami pola distribusi kasus secara mendalam dan mengidentifikasi wilayah-wilayah yang memiliki karakteristik kerawanan serupa, diperlukan sebuah metode analisis yang mampu mengelompokkan data secara objektif. Dalam konteks ini, pengelompokan (*clustering*) menjadi alat yang esensial. Metode *clustering* K-Means dikenal efektif dalam mengidentifikasi kelompok data (*cluster*) dengan karakteristik yang homogen di dalamnya, namun heterogen di antara kelompok-kelompok tersebut[2]. Dengan menerapkan K-Means pada data kekerasan anak per kabupaten/kota di Sumatera Selatan, kita dapat mengelompokkan daerah berdasarkan dominasi jenis kekerasan, sehingga mempermudah pemerintah daerah dan lembaga perlindungan anak dalam merancang kebijakan yang lebih tepat sasaran. Oleh karena itu, penelitian ini bertujuan untuk mengaplikasikan metode K-Means untuk menganalisis dan mengelompokkan kabupaten/kota di Sumatera Selatan berdasarkan data jumlah kekerasan terhadap anak menurut jenis kekerasan yang dialami pada tahun 2023.

Dengan menerapkan metode *clustering* K-Means melalui pemrograman R menggunakan struktur objek S3, penelitian ini berupaya memetakan kabupaten/kota di Provinsi Sumatera Selatan berdasarkan variasi jenis kekerasan anak yang terjadi pada tahun 2023. Pendekatan ini diharapkan mampu mengungkap pola kerawanan yang tidak tampak secara langsung dari data mentah. Hasil klasifikasi wilayah tersebut diharapkan dapat memberikan dasar analitis dalam perencanaan strategi pencegahan serta kebijakan perlindungan anak yang lebih tepat sasaran.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana penerapan algoritma K-Means *Clustering* untuk mengelompokkan kabupaten/kota di Sumatera Selatan berdasarkan jenis kekerasan terhadap anak yang dialami?
2. Bagaimana hasil pengelompokan (*cluster*) kabupaten/kota di Sumatera Selatan berdasarkan dominasi jenis kekerasan (Fisik, Psikis, Seksual, dll.) yang terjadi?

3. Bagaimana interpretasi dari hasil *clustering* K-Means dapat digunakan sebagai rekomendasi untuk penentuan prioritas dan strategi pencegahan kekerasan anak di Sumatera Selatan.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menerapkan algoritma K-Means *Clustering* pada dataset jumlah kekerasan terhadap anak menurut jenis kekerasan di kabupaten/kota Sumatera Selatan tahun 2023.
2. Menghasilkan model pengelompokan (klaster) yang memetakan kabupaten/kota di Sumatera Selatan berdasarkan karakteristik kekerasan anak yang dominan.
3. Memberikan interpretasi dan rekomendasi strategis bagi instansi terkait dalam menyusun kebijakan pencegahan dan penanganan kekerasan anak yang lebih fokus dan efektif sesuai dengan karakteristik klaster wilayah.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi bagi:

1. Pemerintah Daerah (Dinas Pemberdayaan Perempuan dan Perlindungan Anak): Memberikan informasi terstruktur mengenai pola kerawanan kasus kekerasan terhadap anak di setiap wilayah dalam bentuk klaster, sehingga dapat digunakan sebagai dasar dalam menentukan prioritas program, alokasi sumber daya, serta perumusan strategi penanganan yang lebih tepat sasaran.
2. Akademisi: Menjadi referensi dalam pengembangan penelitian terkait penerapan metode *data mining*, khususnya algoritma K-Means dalam konteks kajian sosial dan kesehatan masyarakat, serta memperkaya literatur mengenai analisis spasial dan *clustering* pada isu perlindungan anak.
3. Masyarakat: Meningkatkan pemahaman dan kesadaran masyarakat terhadap kondisi kerawanan kekerasan anak di wilayah sekitarnya, sehingga dapat mendorong partisipasi aktif dalam upaya pencegahan serta penguatan lingkungan yang lebih aman bagi anak.

BAB II

TINJAUAN PUSTAKA

2.1 Kekerasan Terhadap Anak

Menurut Undang-Undang Nomor 35 Tahun 2014 Anak adalah seseorang yang belum berusia 18 (delapan belas) tahun, termasuk anak yang masih dalam kandungan[3]. Undang-Undang Nomor 35 Tahun 2014 Pasal 1 angka 15a tentang Perlindungan Anak menyatakan bahwa kekerasan adalah setiap perbuatan terhadap Anak yang berakibat timbulnya kesengsaraan atau penderitaan secara fisik, psikis, seksual, dan/atau penelantaran, termasuk ancaman untuk melakukan perbuatan, pemaksaan, atau perampasan kemerdekaan secara melawan hukum[3]. Artinya anak yang mengalami kekerasan fisik maupun psikis berdampak pada perilaku di usia-usia berikutnya. Masalah perilaku ini tampak saat anak berada di lingkungan sosial sehingga mengalami sulit berkonsentrasi, malas sekolah, kurang percaya diri, mudah cemas, dan lain-lain[4].

Kekerasan anak adalah perilaku orang dewasa atau anak yang lebih tua dengan menggunakan kekuasaan/otoritasnya terhadap anak yang tak berdaya yang seharusnya menjadi tanggung jawab dari orangtua atau pengasuh yang berakibat penderitaan, kesengsaraan, cacat/kematian. Kekerasan pada anak lebih bersifat sebagai bentuk penganiayaan fisik dengan terdapatnya tanda atau luka pada tubuh sang anak[4].

2.2 Clustering

Clustering merupakan bidang penelitian dalam analisis dan *data mining*. Pada *clustering* partisi dari objek data yang mempunyai karakteristik sama akan dikelompokkan pada satu kelompok dan data yang memiliki karakteristik berbeda akan dikelompokkan pada kelompok yang lainnya. Salah satu metode pada *clustering* yang terkenal adalah K-means[5].

K-means yaitu salah satu dari metode pengelompokkan data non-hierarki yang dapat mempartisi data ke dalam bentuk dua kelompok ataupun lebih. Metode tersebut akan mempartisi data ke dalam suatu kelompok dimana data yang berkarakteristik sama akan dimasukkan ke dalam satu kelompok yang sama, sedangkan data yang memiliki karakteristik berbeda akan dikelompokkan ke dalam kelompok lainnya[5]. Rumus k-means:

$$J = \sum_{i=1}^k \sum ||x - \mu_i||^2$$

Keterangan :

J : Total Sum of Squares

k : Jumlah Klaster

x : Titik data

μ_i : Pusat Klaster i

Algoritma K-Means merupakan salah satu teknik pengelompokan non-hirarki dalam *data clustering* yang bekerja dengan membagi sekumpulan data ke dalam beberapa klaster berdasarkan kemiripan atributnya. Pada metode ini, data yang memiliki karakteristik serupa akan ditempatkan dalam kelompok yang sama, sedangkan data yang berbeda akan dipisahkan ke dalam kelompok lain. Tujuan penerapan K-Means adalah meminimalkan nilai fungsi objektif selama proses pengelompokan, yang umumnya dicapai dengan mengurangi variasi di dalam klaster dan meningkatkan perbedaan antar-klaster sehingga hasil pengelompokan menjadi lebih terstruktur dan informatif[6].

2.3 Principal Component Analysis (PCA)

Metode *Principal Component Analysis* (PCA) merupakan suatu teknik multivariat. Secara umum, metode PCA digunakan untuk mereduksi faktor atau variabel dalam jumlah besar menjadi beberapa faktor yang lebih sedikit tanpa perlu menghilangkan informasi penting yang terkandung dalam data. Seringkali informasi tersebut dikelompokkan menjadi variabel baru yang lebih sederhana berdasarkan kemiripan informasi yang diperoleh dari data awal[7].

2.4 Pemrograman Fungsi R Class S3

Pemrograman fungsi di R merupakan dasar dalam proses analisis data karena hampir seluruh operasi dilakukan melalui fungsi. Salah satu fitur penting dalam R adalah dukungan terhadap *object-oriented programming* (OOP) melalui sistem Class S3, yaitu sistem objek yang sederhana dan fleksibel. Sesuai penjelasan Tiwari (2014), S3 bekerja dengan menambahkan atribut class pada objek sehingga objek tersebut dapat dikenali sebagai jenis tertentu[8].

S3 memanfaatkan *generic functions* seperti `print()`, `summary()`, dan `plot()`, yang akan memanggil metode khusus berdasarkan kelas objek (*method dispatch*). Programmer dapat membuat fungsi generik baru menggunakan `UseMethod()` dan mendefinisikan metode untuk kelas tertentu melalui penamaan `generic.class`. Pendekatan ini memungkinkan fungsi yang sama berperilaku berbeda untuk jenis objek yang berbeda[8].

Dengan strukturnya yang sederhana, tidak ketat, dan mudah diperluas, Class S3 menjadi sistem OOP yang paling banyak digunakan dalam R. Pemahaman mengenai pemrograman fungsi dan mekanisme Class S3 membantu pengguna menghasilkan kode yang lebih rapi, modular, dan mudah dipelihara[8].

BAB III

METODOLOGI PENELITIAN

3.1 Jenis Data

Penelitian ini menggunakan pendekatan deskriptif kuantitatif untuk menganalisis kasus kekerasan terhadap anak berdasarkan jenis kekerasan yang terjadi di Provinsi Sumatera Selatan pada tahun 2023. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari publikasi resmi Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KemenPPPA). Data tersebut disajikan dalam bentuk laporan statistik mengenai jumlah kekerasan terhadap anak menurut jenis kekerasan yang dialami pada masing-masing kabupaten/kota.

3.2 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan menggunakan metode studi dokumentasi. Metode ini digunakan karena penelitian sepenuhnya memanfaatkan data sekunder yang telah disediakan oleh lembaga resmi pemerintah. Data yang digunakan bersumber dari Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KemenPPPA), berupa publikasi statistik mengenai jumlah kasus kekerasan terhadap anak menurut jenis kekerasan yang dialami di Provinsi Sumatera Selatan pada tahun 2023.

Pengumpulan data dilakukan dengan cara mengunduh dan menelaah dokumen resmi tersebut, kemudian mengidentifikasi variabel-variabel yang relevan untuk dianalisis. Selanjutnya, data diolah dan disajikan kembali dalam bentuk tabel dan visualisasi untuk mendukung proses analisis deskriptif dalam penelitian. Metode ini dipilih karena data yang diperlukan telah tersedia secara lengkap dan resmi.

3.3 Variabel yang Diamati

Dalam penelitian ini digunakan sejumlah variabel yang berkaitan dengan jenis kekerasan terhadap anak di setiap kabupaten/kota di Provinsi Sumatera Selatan. Variabel-variabel tersebut menjadi dasar dalam proses analisis dan pengelompokan menggunakan metode *clustering* K-Means. Daftar lengkap variabel yang digunakan disajikan pada Tabel 3.1, sedangkan dataset penelitian yang menjadi sumber data analisis ditampilkan pada Tabel 3.2.

Dalam penelitian ini, setiap variabel dirumuskan secara operasional untuk memastikan bahwa proses pengukuran berlangsung secara objektif dan konsisten. Operasionalisasi variabel dilakukan dengan menetapkan definisi teknis, satuan pengukuran, serta skala data yang digunakan sehingga setiap variabel dapat dianalisis secara kuantitatif. Tabel 3.1 berikut menyajikan daftar variabel yang digunakan dalam penelitian beserta definisi operasionalnya sebagai dasar dalam proses analisis *clustering* K-Means. Selain itu, perumusan variabel ini juga berperan penting dalam menjaga kesesuaian antara tujuan penelitian dan metode analisis yang diterapkan.

Tabel 3.1 Operasionalisasi Variabel

Variabel	Definisi	Satuan	Skala
Kekerasan Fisik	Jumlah kasus kekerasan berupa tindakan fisik yang dialami anak, sebagaimana tercatat dalam data KemenPPPA Tahun 2023.	Jumlah Kasus	Rasio
Kekerasan Psikis	Jumlah kasus kekerasan yang memberikan dampak tekanan mental atau emosional terhadap anak.	Jumlah Kasus	Rasio
Kekerasan Seksual	Jumlah kasus kekerasan seksual yang dialami anak, termasuk pelecehan dan eksploitasi seksual.	Jumlah Kasus	Rasio
Eksploitasi	Jumlah kasus tindakan eksploitasi terhadap anak yang tercatat dalam laporan KemenPPPA.	Jumlah Kasus	Rasio
TPPO (Tindak Pidana Perdagangan Orang)	Jumlah kasus perdagangan anak untuk tujuan eksploitasi menurut data resmi KemenPPPA Tahun 2023.	Jumlah Kasus	Rasio
Penelantaran	Jumlah kasus anak yang tidak mendapatkan pemenuhan kebutuhan dasar maupun pengasuhan yang layak.	Jumlah Kasus	Rasio
Kekerasan Lainnya	Jumlah kasus kekerasan lain yang tidak termasuk dalam kategori utama tetapi tercatat dalam laporan KemenPPPA.	Jumlah Kasus	Rasio
Wilayah Administratif (Kabupaten/Kota)	Unit wilayah pemerintahan di Provinsi Sumatera Selatan yang menjadi lokasi terjadinya kasus.	Nama Wilayah	Kategorik
Total Kasus Kekerasan	Jumlah seluruh kasus kekerasan terhadap anak pada suatu kabupaten/kota.	Jumlah Kasus	Rasio

Tabel 3.2 menyajikan data penelitian yang menjadi dasar dalam proses analisis serta pengelompokan wilayah berdasarkan kategori kekerasan terhadap anak di Provinsi Sumatera Selatan. Data tersebut memuat jumlah kasus pada setiap kabupaten/kota yang diperoleh dari sumber resmi KemenPPPA. Penyajian tabel ini bertujuan memberikan gambaran awal

mengenai persebaran kasus sehingga membantu peneliti memahami konteks data sebelum melangkah ke tahap analisis yang lebih mendalam.

Selain itu, informasi dalam Tabel 3.2 berfungsi untuk menunjukkan variasi jumlah kasus di setiap daerah, sehingga peneliti dapat mengidentifikasi wilayah dengan tingkat kejadian yang lebih tinggi ataupun lebih rendah. Pemahaman awal ini sangat penting agar tahapan analisis selanjutnya termasuk proses normalisasi data dan penetapan jumlah kluster dapat dilakukan sesuai dengan karakteristik data sebenarnya. Oleh karena itu, ketelitian data pada tabel ini turut menentukan kualitas hasil pengelompokan yang dihasilkan.

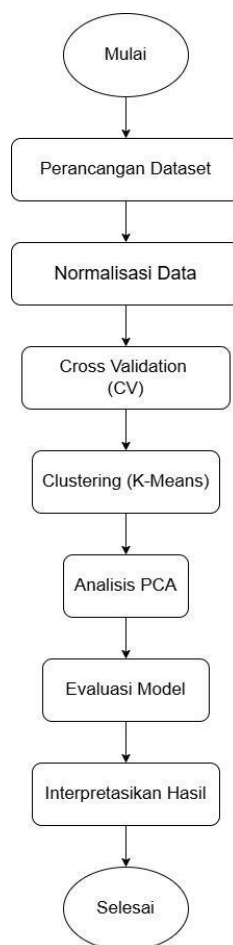
Tabel 3.2 Dataset Penelitian

Kabupaten / Kota	Kekerasan			Eksplorasi	TPPO (Tindak Pidana Perdagangan Orang)	Penelantaran	Kekerasan Lainnya
	Fisik	Psikis	Seksual				
Ogan Komering Ulu	4	4	3	0	0	0	0
Ogan Komering Ilir	34	35	15	0	0	1	2
Muara Enim	23	20	46	0	1	1	6
Lahat	1	5	15	0	0	5	2
Musi Rawas	2	0	0	0	0	0	1
Musi Banyuasin	0	0	0	0	0	0	0
Banyu Asin	5	1	26	1	0	0	4
Ogan Komering Ulu Selatan	7	0	11	0	0	0	2
Ogan Komering Ulu Timur	14	6	36	0	0	0	2
Ogan Ilir	28	0	17	0	0	0	2
Empat Lawang	0	0	0	0	1	4	28
Penukal Abab Lematang Ilir	4	3	10	0	0	0	5
Musi Rawas Utara	0	0	16	0	0	0	0
Palembang	11	13	35	1	0	1	13

Kabupaten / Kota	Kekerasan			Eksplorasi	TPPO (Tindak Pidana Perdagangan Orang)	Penelantaran	Kekerasan Lainnya
	Fisik	Psikis	Seksual				
Prabumulih	2	2	20	0	0	0	0
Pagar Alam	35	3	13	0	0	0	4
Lubuklinggau	2	9	11	0	1	0	0

3.4 Diagram Alir

Pada bagian ini disajikan diagram alir yang menggambarkan tahapan-tahapan utama dalam proses analisis data yang dilakukan dalam penelitian ini. Diagram tersebut berfungsi untuk memberikan gambaran menyeluruh mengenai alur kerja mulai dari tahap persiapan data hingga proses evaluasi dan interpretasi hasil. Dengan adanya diagram alir ini, pembaca dapat memahami struktur dan urutan prosedur analisis secara lebih jelas, sehingga memudahkan dalam mengikuti setiap langkah yang dilakukan peneliti selama proses pengolahan data.



Gambar 3.1 Diagram Alir Proses Analisis Data dalam Penelitian

Proses analisis data pada penelitian ini diawali dengan tahap perancangan dan pengumpulan dataset, yaitu mengumpulkan data kasus kekerasan anak dari publikasi resmi pada tingkat kabupaten/kota di Provinsi Sumatera Selatan. Setelah dataset berhasil diimpor, dilakukan proses pembersihan dan penstrukturan data, meliputi pemeriksaan kesesuaian format variabel, konsistensi nilai, serta penanganan nilai kosong atau tidak valid agar siap digunakan dalam analisis. Tahap berikutnya adalah normalisasi data. Normalisasi diperlukan karena setiap jenis kekerasan seperti kekerasan fisik, psikis, seksual, eksploitasi, dan penelantaran memiliki rentang nilai yang berbeda. Dengan normalisasi, setiap variabel memiliki kontribusi yang seimbang sehingga tidak ada variabel yang mendominasi proses pembentukan kluster pada algoritma K-Means.

Setelah data dinormalisasi, dilakukan *Cross Validation* (CV), yang digunakan untuk menilai stabilitas model dalam memilih jumlah kluster yang paling optimal. Tahap ini mengevaluasi bagaimana variasi hasil kluster berubah terhadap pembagian data secara berulang, sehingga pemilihan parameter K menjadi lebih objektif dan tidak bias. Selanjutnya dilakukan *Clustering* menggunakan algoritma K-Means, dengan memasukkan variabel numerik yang terkait jenis kekerasan anak sebagai input. Proses ini mengelompokkan kabupaten/kota berdasarkan pola kesamaan kasus kekerasan. Hasil *clustering* disimpan dengan menggunakan mekanisme S3 Class di R, sehingga hasil analisis tersimpan dalam struktur data yang lebih sistematis, dan memudahkan penggunaan fungsi evaluasi maupun visualisasi.

Untuk mendukung interpretasi hasil kluster, dilakukan PCA (*Principal Component Analysis*). PCA mereduksi dimensi data sehingga pola antarwilayah dapat divisualisasikan dalam dua komponen utama. Representasi visual PCA memudahkan peneliti untuk mengidentifikasi kedekatan antarwilayah dan perbedaan karakteristik pada tiap kluster.

Tahap selanjutnya adalah evaluasi model, yaitu dengan memeriksa pemisahan kluster, nilai *within-cluster sum of squares* (WCSS), serta validasi visual seperti *scatter plot* dan *heatmap*. Evaluasi ini memastikan bahwa kluster yang terbentuk memiliki perbedaan yang cukup jelas antar kelompok wilayah sehingga tiap kluster benar-benar merepresentasikan karakteristik pola kasus yang berbeda. Proses analisis ini diakhiri dengan tahap interpretasi hasil dan penyimpulan sebagai penutup dari keseluruhan rangkaian analisis.

BAB IV

HASIL DAN PEMBAHASAN

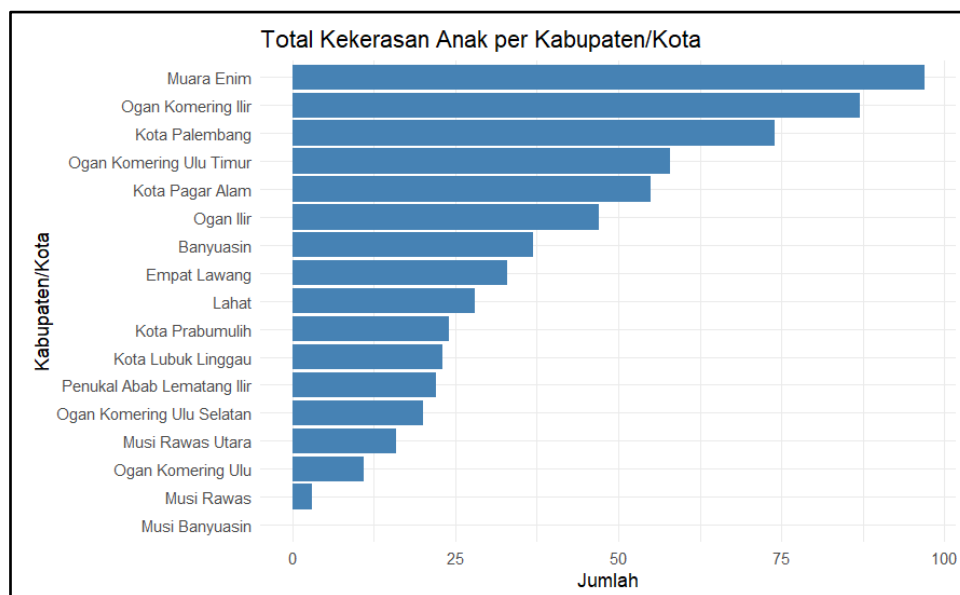
4.1 Deskripsi Data

Data yang digunakan dalam penelitian ini merupakan data jumlah kasus kekerasan terhadap anak pada tingkat kabupaten/kota di Provinsi Sumatera Selatan tahun 2023, yang terdiri dari 17 observasi mewakili seluruh wilayah administrasi provinsi tersebut. Variabel yang dicatat mencakup berbagai jenis kekerasan, antara lain kekerasan fisik, psikis, seksual, eksploitasi, Tindak Pidana Perdagangan Orang (TPPO), penelantaran, serta kategori lainnya. Secara umum, distribusi kasus menunjukkan bahwa kekerasan fisik, psikis, dan seksual merupakan tiga kategori yang paling dominan dilaporkan oleh masing-masing daerah, dengan jumlah kasus yang bervariasi cukup besar antar kabupaten/kota. Beberapa daerah seperti Kabupaten Ogan Komering Ilir, Kabupaten Lahat, dan Kabupaten Musi Rawas Utara terlihat mencatat jumlah kasus seksual yang relatif tinggi, sedangkan variabel eksploitasi dan TPPO cenderung memiliki nilai yang sangat rendah, bahkan tidak tercatat pada sebagian besar wilayah. Kategori penelantaran juga menunjukkan variasi kecil dengan jumlah kasus maksimal sebanyak lima kasus, sementara kategori lainnya mencatat variasi yang lebih besar hingga 28 kasus pada wilayah tertentu. Secara keseluruhan, dataset ini memberikan gambaran mengenai persebaran jenis-jenis kekerasan terhadap anak di Sumatera Selatan, dengan pola yang mengindikasikan bahwa kekerasan fisik, psikis, dan seksual merupakan permasalahan yang paling menonjol pada tahun pengamatan.

Tabel 4.1 Summary Dataset

Variabel	Mean	Median	Mode	Min	Max	SD (Standar Deviasi)	Variance	IQR
Fisik	10.11764 71	4	Numeric	0	35	12.22130 49	149.3602 941	12
Psikis	5.941176 5	3	Numeric	0	35	9.270049 8	85.93382 35	6
Seksual	16.11764 71	15	Numeric	0	46	13.25651 89	175.7352 941	10
Eksploita si	0.117647 1	0	Numeric	0	1	0.332105 6	0.110294 1	0
TPPO	0.176470 6	0	Numeric	0	1	0.392952 6	0.154411 8	0
Penelanta ran	0.705882 4	0	Numeric	0	5	1.490163 8	2.220588 2	1
Lainnya	4.176470 6	2	Numeric	0	28	6.939338 0	48.15441 18	4

Berdasarkan hasil analisis statistik deskriptif pada tujuh variabel kasus kekerasan terhadap anak di Provinsi Sumatera Selatan, terlihat bahwa jenis kekerasan Seksual memiliki rata-rata (mean) tertinggi yaitu 16,12 kasus, dengan median 15 dan rentang antar-kuartil (IQR) 10, menunjukkan penyebaran data yang cukup besar. Variabel Fisik menempati urutan berikutnya dengan mean 10,12 dan IQR 12, mengindikasikan variasi kasus yang relatif tinggi antarwilayah. Untuk kekerasan Psikis, nilai mean sebesar 5,94 dan median 3, dengan standar deviasi cukup besar (9,27), mengisyaratkan adanya ketimpangan distribusi kasus di beberapa kabupaten/kota. Sementara itu, variabel Eksploitasi dan TPPO menunjukkan nilai mean sangat rendah (masing-masing 0,12 dan 0,18), dengan median 0, menandakan bahwa sebagian besar wilayah tidak mengalami kasus tersebut pada tahun pengamatan. Jenis kekerasan Penelantaran memiliki rata-rata 0,71 dengan variasi sedang, sedangkan kategori Lainnya mencatat mean 4,18 dan IQR 4, menunjukkan adanya jumlah kasus tambahan yang cukup bervariasi. Secara keseluruhan, data ini menunjukkan bahwa kasus kekerasan seksual dan fisik mendominasi, sementara jenis kekerasan lainnya relatif jarang terjadi.



Gambar 4.2 Visualisasi Data Menggunakan Bar Chart

Berdasarkan Gambar 4.2, ditampilkan total kasus kekerasan terhadap anak pada setiap kabupaten/kota di Provinsi Sumatera Selatan. Diagram batang tersebut menunjukkan adanya perbedaan yang cukup signifikan antarwilayah dalam jumlah kasus yang dilaporkan. Beberapa daerah, seperti Kabupaten Muara Enim, Kabupaten Ogan Komering Ilir, dan Kota Palembang, terlihat memiliki jumlah kasus yang relatif tinggi, sedangkan beberapa wilayah lainnya mencatat angka yang jauh lebih rendah.

Variasi ini menunjukkan perlunya perhatian yang berbeda pada setiap daerah serta pentingnya analisis lanjutan untuk memahami faktor-faktor yang memengaruhi tingginya angka kekerasan di wilayah tertentu. Selain itu, visualisasi ini dapat menjadi dasar awal dalam menentukan prioritas intervensi dan perumusan kebijakan perlindungan anak yang lebih efektif di masing-masing wilayah. Dengan demikian, hasil visualisasi ini dapat membantu dalam mengambil keputusan berbasis data secara lebih tepat.

4.2 Hasil *Clustering*

Hasil pada tabel 4.1 menyajikan perhitungan jarak setiap kabupaten/kota terhadap pusat klaster (centroid) yang terbentuk dari proses klasterisasi. Nilai jarak ini digunakan untuk menunjukkan kedekatan masing-masing wilayah dengan klaster terpilih, sehingga dapat diketahui tingkat kemiripan karakteristik antarwilayah. Semakin kecil nilai jarak terhadap suatu centroid, maka semakin kuat wilayah tersebut merepresentasikan klaster yang bersangkutan.

Tabel 4.1 Jarak Kabupaten terhadap Pusat Klaster (Centroid)

Kabupaten	Cluster Terpilih	Jarak_ke_C1	Jarak_ke_C2	Jarak_ke_C3
Ogan Komering Ulu	1	8.518.332	3.207.976	39.631.637
Ogan Komering Ilir	2	45.812.049	2.241.527	37.541.089
Muara Enim	3	45.033.505	333.833	12.192.894
Lahat	1	7.355.776	3.260.027	30.149.627
Musi Rawas	1	10.830.016	3.617.243	43.848.223
Musi Banyuasin	1	11.390.514	3.792.244	44.672.885
Banyuasin	1	16.113.749	3.173.501	21.071.308
Ogan Komering Ulu Selatan	1	5.487.025	2.861.429	32.552.522
Ogan Komering Ulu Timur	3	28.608.933	2.867.248	9.363.048
Ogan Ilir	2	26.605.163	1.355.647	28.682.166
Empat Lawang	1	26.646.135	4.568.491	48.976.185
Penukal Abab Lematang Ilir	1	2.279.988	3.044.303	33.010.099
Musi Rawas Utara	1	7.741.164	348.441	31.680.699
Kota Palembang	3	29.852.945	310.394	8.812.869
Kota Prabumulih	1	10.579.491	3.265.136	26.975.297
Kota Pagar Alam	2	32.688.866	1.031.719	33.862.467
Kota Lubuk Linggau	1	7.955.454	3.094.799	32.341.923

Hasil analisis klaster terhadap data kabupaten/kota di Provinsi Sumatera Selatan menunjukkan pembagian wilayah menjadi tiga kelompok utama berdasarkan kedekatan jarak masing-masing kabupaten terhadap *centroid* klaster. *Cluster 1* merupakan kelompok dengan jumlah anggota terbanyak, yang berisi mayoritas kabupaten seperti Ogan Komering Ulu, Lahat, Musi Rawas, Musi Banyuasin, Banyuasin, serta beberapa kota seperti Prabumulih dan Lubuk Linggau. Daerah-daerah ini memiliki karakteristik indikator yang relatif dekat satu sama lain sehingga jaraknya terhadap centroid *Cluster 1* menjadi yang paling kecil. *Cluster 2* beranggotakan wilayah yang memiliki pola nilai berbeda, khususnya Ogan Komering Ilir, Ogan Ilir, dan Kota Pagar Alam, yang ketiganya menunjukkan jarak minimum terhadap centroid *Cluster 2*

dibandingkan *centroid* lainnya. Sementara itu, *Cluster 3* terdiri dari wilayah dengan karakteristik paling berbeda dari kedua klaster sebelumnya, yaitu Muara Enim, Ogan Komering Ulu Timur, dan Kota Palembang, yang secara konsisten menunjukkan jarak terdekat ke *centroid Cluster 3*. Pembagian klaster ini memberikan gambaran bahwa masing-masing kelompok wilayah memiliki kesamaan internal yang kuat, namun berbeda secara signifikan antar klaster, sehingga dapat menjadi dasar untuk analisis lanjutan atau pengambilan keputusan berbasis karakteristik wilayah.

Tabel 4.2 Rata-Rata Jenis Kekerasan dan Kategori Kasus per Klaster

<i>Cluster</i>	Fisik	Psikis	Seksual	Eksplorasi	TPPO	Penelantaran	Lainnya	Total_Skor	Kategori
1	2.454545	2.181818	1.018182	9.090909	1.818	8.181818	3.818182	1.972727	Kasus Rendah
2	32.333333	12.666667	15	0	0	3.333333	2.666667	63	Kasus Sedang
3	16	13	39	0.333333	0.333333	0.666667	7	76.33333	Kasus Tinggi

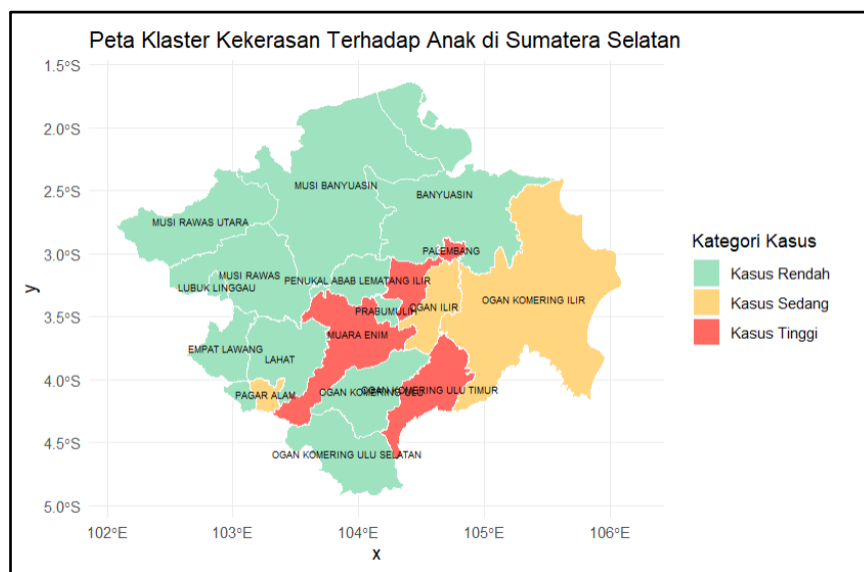
Hasil analisis klaster terhadap data kekerasan anak menghasilkan tiga kelompok wilayah yang memiliki karakteristik tingkat kasus yang berbeda. *Cluster 1*, yang dikategorikan sebagai Kasus Rendah, menunjukkan rata-rata insiden kekerasan yang relatif kecil di hampir semua jenis kasus. Kekerasan fisik dan psikis berada pada kisaran 2 kasus, sedangkan kekerasan seksual sekitar 10 kasus. Nilai untuk kategori eksploitasi, TPPO, dan penelantaran juga sangat rendah, masing-masing kurang dari 1 kasus rata-rata, dengan total skor 19,72. Hal ini mencerminkan bahwa wilayah dalam klaster ini memiliki tingkat kekerasan yang masih dapat dikatakan terkendali. Berbeda dengan itu, *Cluster 2* masuk kategori Kasus Sedang, ditandai dengan tingginya kasus kekerasan fisik (32,33 kasus) dan psikis (12,67 kasus). Kekerasan seksual juga berada pada tingkat menengah dengan 15 kasus rata-rata. Meskipun eksploitasi dan TPPO tidak muncul pada klaster ini, kasus penelantaran dan kategori lainnya tetap ada, menghasilkan total skor 63. Wilayah dalam klaster ini memiliki tingkat kekerasan yang lebih serius dibandingkan klaster rendah, namun belum mencapai kondisi paling ekstrem.

Sementara itu, *Cluster 3* merupakan kelompok dengan tingkat kekerasan paling tinggi dan masuk kategori Kasus Tinggi. Klaster ini memiliki rata-rata kasus seksual yang sangat mencolok, yaitu 39 kasus, serta kekerasan fisik dan psikis yang tinggi masing-masing 16 dan 13 kasus. Meskipun eksploitasi, TPPO, dan penelantaran hanya muncul dalam jumlah kecil, dominasi kasus seksual dan jenis kekerasan lainnya mengangkat total skor klaster ini menjadi 76,33 yang tertinggi dari seluruh kelompok. Kondisi ini menggambarkan bahwa wilayah-wilayah dalam *Cluster 3* menghadapi permasalahan kekerasan anak yang jauh lebih berat dan membutuhkan perhatian serta intervensi yang lebih intensif.

4.3 Visualisasi

Pada tahap visualisasi, analisis *clustering* diperkuat melalui penyajian grafis untuk memperjelas pola hubungan antarwilayah serta karakteristik masing-masing klaster.

Visualisasi berperan sebagai alat bantu dalam menafsirkan hasil pengelompokan yang dihasilkan oleh algoritma K-Means, sehingga perbedaan tingkat kasus antar kabupaten/kota dapat diamati secara lebih intuitif. Melalui pendekatan *Principal Component Analysis* (PCA), data berdimensi tinggi direduksi menjadi dua komponen utama yang mampu menangkap proporsi variasi terbesar dalam dataset. Hasil reduksi tersebut kemudian divisualisasikan dalam bentuk peta kluster, sehingga setiap wilayah tidak hanya terlihat berdasarkan jarak numerik antar data, tetapi juga dalam konteks spasial dan persebaran geografisnya. Visualisasi ini menjadi langkah penting dalam memahami kedekatan karakteristik antarwilayah, mengidentifikasi pola kerawanan, serta menegaskan perbedaan yang muncul antar kluster secara lebih tepat. Dengan demikian, peta kluster memberikan gambaran komprehensif yang mendukung proses analisis dan interpretasi terhadap sebaran kasus kekerasan anak.



Gambar 4.3 Visualisasi Menggunakan PCA

Peta kluster kekerasan terhadap anak di Provinsi Sumatera Selatan menampilkan hasil pengelompokan wilayah berdasarkan tingkat kasus menggunakan metode K-Means. Setiap kabupaten/kota diberi warna sesuai kategori kluster, yaitu hijau untuk Kasus Rendah, oranye untuk Kasus Sedang, dan merah untuk Kasus Tinggi. Pola warna pada peta menunjukkan bahwa daerah dengan risiko tinggi banyak terkonsentrasi di bagian tengah dan selatan provinsi, sedangkan wilayah di utara dan barat didominasi kategori rendah. Label nama kabupaten/kota ditempatkan pada masing-masing centroid sehingga memudahkan identifikasi lokasi. Peta ini memberikan gambaran spasial yang jelas mengenai distribusi tingkat kekerasan anak, sehingga mempermudah analisis wilayah yang perlu perhatian prioritas dalam penanganan kasus.

4.4 Diskusi

Berdasarkan perbandingan dengan laporan terdahulu berjudul “*Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritma K-Means dengan Perbandingan Jarak Manhattan City dan Euclidean*”, terdapat beberapa perbedaan penting dalam pendekatan metodologis maupun keluaran analisis. Pada penelitian tersebut, fokus utama terletak pada perbandingan dua metrik jarak, yaitu *Euclidean Distance* dan *Manhattan City Block*, untuk menilai pengaruh perbedaan

jarak terhadap hasil pengelompokan kasus kekerasan anak. Analisis mereka berorientasi pada evaluasi performa algoritma berdasarkan nilai SSE (*Sum of Squared Error*) sehingga penekanannya ada pada optimasi metode dan perbandingan akurasi antarjarak. Sementara itu, penelitian yang dilakukan dalam laporan ini lebih menekankan pada pemrosesan struktural dan modular melalui pemrograman fungsi berbasis Class S3 di R, sehingga penelitian tidak hanya menghasilkan klaster, tetapi juga menghasilkan framework analisis yang sistematis, dapat diperluas, dan terotomatisasi. Selain itu, penelitian ini memperkaya analisis dengan beberapa teknik visualisasi lanjutan seperti heatmap, PCA (*Principal Component Analysis*), serta peta spasial, yang tidak dijumpai dalam laporan terdahulu. Jika penelitian sebelumnya hanya menilai apakah klaster terbentuk dengan baik berdasarkan nilai SSE dan jarak terpilih, penelitian ini memberikan pendekatan interpretatif yang lebih komprehensif dengan menilai karakteristik masing-masing klaster berdasarkan rata-rata jenis kasus serta memetakan klaster tersebut secara spasial di tingkat kabupaten/kota. Dengan demikian, penelitian ini bukan hanya membandingkan performa algoritma, tetapi juga menghasilkan insight substantif mengenai pola kerawanan wilayah, sehingga lebih relevan sebagai dasar perumusan kebijakan mitigasi dan intervensi kekerasan terhadap anak.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini menunjukkan bahwa analisis pengelompokan menggunakan algoritma K-Means mampu mengidentifikasi pola persebaran kasus kekerasan anak di Provinsi Sumatera Selatan secara jelas. Tiga klaster yang terbentuk menggambarkan perbedaan tingkat kerawanan antarwilayah, di mana beberapa kabupaten/kota berada pada klaster berisiko tinggi terutama pada kategori kekerasan seksual dan psikis. Implementasi metode melalui pendekatan pemrograman R berbasis S3 juga terbukti efektif dan fleksibel dalam mengelola data, melakukan visualisasi, serta membangun fungsi-fungsi analitis seperti heatmap, PCA, dan pie chart. Temuan ini menegaskan bahwa pengolahan data terstruktur dengan pendekatan statistik dan pemrograman memberikan kontribusi penting dalam memahami dinamika permasalahan kekerasan anak, sehingga dapat digunakan sebagai dasar pengambilan keputusan maupun penyusunan strategi intervensi.

5.2 Saran

Penelitian ini dapat ditingkatkan pada beberapa aspek penting. Pertama, akan lebih baik jika penentuan jumlah *cluster* tidak hanya ditetapkan langsung menjadi tiga, tetapi juga diuji menggunakan metode evaluasi seperti Elbow Method atau Silhouette Coefficient agar struktur klaster yang dipilih benar-benar optimal. Selain itu, proses standarisasi data juga sebaiknya diterapkan sebelum melakukan pengelompokan, mengingat setiap variabel memiliki satuan dan skala yang berbeda sehingga dapat memengaruhi pembentukan centroid. Pengayaan analisis visual seperti menambahkan grafik radar untuk profil masing-masing *cluster* juga dapat membantu dalam memahami perbedaan karakteristik antara kelompok wilayah. Kemudian, hasil klaster dapat dipadukan dengan informasi eksternal misalnya jumlah penduduk anak di tiap kabupaten agar interpretasi tingkat keparahan kasus tidak bias oleh perbedaan ukuran wilayah. Jika memungkinkan, penggunaan metode *clustering* lain seperti *Hierarchical Clustering* atau *Gaussian Mixture Model* dapat dijadikan pembanding untuk memastikan konsistensi hasil analisis. Dengan beberapa penyempurnaan tersebut, penelitian akan memiliki landasan analitis yang lebih kuat dan interpretasi yang lebih komprehensif terhadap fenomena kekerasan terhadap anak di Sumatera Selatan.

DAFTAR PUSTAKA

- [1] Putu Arga, Muhammad. (2022). PERAN UNITED NATION CHILDREN'S FUND (UNICEF) DALAM MENGATASI KEKERASAN ANAK DAN PENJUALAN ANAK DI ASIA.
- [2] Buaton, Relita & Sundari, Yeni & Maulita, Yani. (2016). Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritma K-Means Dengan Perbandingan Jarak Kedekatan Manhattan City Dan Euclidean. *MEANS (Media Informasi Analisa dan Sistem)*. 47-53. 10.54367/means.v1i2.8.
- [3] R. Indonesia, Undang-Undang Nomor 35 Tahun 2014 Pasal 1 angka 15a tentang Perlindungan Anak , Jakarta, 2014.
- [4] W. dan I. Pusnita, "Efektivitas Pelayanan Penanggulangan Kekerasan Anak pada Pusat Pelayanan Terpadu Pemberdayaan Perempuan dan Anak (P2TP2A) Kota Palembang," *Journal of Social Sciences and Politics (PUBLISITAS)*, vol. 9, no. 1, pp. 65–74, Okt. 2022.
- [5] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *JURNAL NASIONAL TEKNOK DAN SISTEM INFORMASI*, vol. 05, no. 01, pp. 017–024, Nov. 2019.
- [6] R. Buaton, Y. Sundari, and Y. Maulita, "Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritma K-Means Dengan Perbandingan Jarak Kedekatan Manhattan City Dan Euclidean," *MEANS (Media Informasi Analisa dan Sistem)*, vol. 1, no. 2, pp. 47-53, Dec. 2016.
- [7] D. R. P. Sari, "Metode Principal Component Analysis (PCA) sebagai Penanganan Asumsi Multikolinearitas (Studi Kasus: Data Produksi Tapioka)," *Parameter: Jurnal Matematika, Statistika dan Terapannya*, vol. 2, no. 2, pp. 115–124, Oct. 2023.
- [8] A. Tiwari, "Class and Objects in R: S3 Style," 2014.

LAMPIRAN

1. Summary Dataset

```
{r}
library(readxl)
library(dplyr)

# Baca file Excel
df <- data

# Ambil hanya kolom numerik
numeric_df <- df %>%
  select(where(is.numeric)) %>%
  select(-No) # kolom No dihapus

# Summary tabel
summary_table <- data.frame(
  Variable = colnames(numeric_df),
  Mean = sapply(numeric_df, function(x) mean(x, na.rm = TRUE)),
  Median = sapply(numeric_df, function(x) median(x, na.rm = TRUE)),
  Mode = sapply(numeric_df, function(x) mode(x)),
  Min = sapply(numeric_df, function(x) min(x, na.rm = TRUE)),
  Max = sapply(numeric_df, function(x) max(x, na.rm = TRUE)),
  SD = sapply(numeric_df, function(x) sd(x, na.rm = TRUE)),
  Variance = sapply(numeric_df, function(x) var(x, na.rm = TRUE)),
  IQR = sapply(numeric_df, function(x) IQR(x, na.rm = TRUE))
)

# Hasil
summary_table
```

2. Visualisasi Barchart dengan Menerapkan Metode S3

```
{r}
visKekerasan <- function(data){
  structure(
    list(data_raw = data),
    class = "visKekerasan"
  )
}

plotVis <- function(obj) {
  UseMethod("plotVis")
}

plotVis.visKekerasan <- function(obj){

  df <- obj$data_raw %>%
    mutate(Total = Fisik + Psikis + Seksual + Eksploitasi + TPPD +
      Penelantaran + Lainnya)

  ggplot(df, aes(x = reorder(Cakupan, Total), y = Total)) +
    geom_col(fill = "steelblue") +
    coord_flip() +
    labs(
      title = "Total Kekerasan Anak per Kabupaten/Kota",
      x = "Kabupaten/Kota",
      y = "Jumlah"
    ) +
    theme_minimal()
}
```

3. Melakukan K-Means *Clustering* dengan Menerapkan Method S3

```
{r}
clustKekerasan <- function(data, k = 3){
  num_data <- data %>%
    select(Fisik, Psikis, Seksual, Eksploitasi, TPPO, Penelantaran, Lainnya)

  set.seed(123)
  km <- kmeans(num_data, centers = k)

  data$Cluster <- km$cluster

  cluster_mean <- aggregate(num_data, list(Cluster = km$cluster), mean)

  # Label otomatis berdasarkan ranking
  ranking <- order(rowSums(cluster_mean[,2:8]))
  cluster_mean$Label <- ""
  cluster_mean$Label[ranking[1]] <- "Kasus Rendah"
  cluster_mean$Label[ranking[2]] <- "Kasus Sedang"
  cluster_mean$Label[ranking[3]] <- "Kasus Tinggi"

  data$Cluster_Label <- cluster_mean$Label[data$Cluster]

  structure(
    list(
      data = data,
      km = km,
      cluster_mean = cluster_mean,
      data_matrix = as.matrix(num_data) %>% `rownames<-`(data$Cakupan),
      cluster = km$cluster
    ),
    class = "clustKekerasan"
  )
}
```

4. Jarak Kabupaten terhadap Pusat *Cluster* (*Centroid*)

```
{r}
# Ambil matrix data dari objek cluster S3
data_matrix <- obj$data_matrix

# Ambil hasil kmeans dari objek
km <- obj$km

# Ambil centroid
centroids <- km$centers

# Fungsi jarak Euclidean
euclid <- function(x, y) sqrt(sum((x - y)^2))

# Buat tabel penjelasan lengkap
penjelasan <- data_matrix %>%
  as.data.frame() %>%
  mutate(
    Kabupaten = rownames(data_matrix),
    Cluster_Terpilih = km$cluster,

    # Hitung jarak ke masing-masing centroid
    Jarak_ke_C1 = apply(data_matrix, 1, function(r) euclid(r, centroids[1,])),
    Jarak_ke_C2 = apply(data_matrix, 1, function(r) euclid(r, centroids[2,])),
    Jarak_ke_C3 = apply(data_matrix, 1, function(r) euclid(r, centroids[3,])),

    # Penjelasan kenapa memilih cluster tertentu
    Alasan = case_when(
      Cluster_Terpilih == 1 ~ "Masuk Cluster 1 karena jarak ke centroid 1 paling kecil",
      Cluster_Terpilih == 2 ~ "Masuk Cluster 2 karena jarak ke centroid 2 paling kecil",
      Cluster_Terpilih == 3 ~ "Masuk Cluster 3 karena jarak ke centroid 3 paling kecil"
    )
  ) %>%
  select(Kabupaten, Cluster_Terpilih,
         Jarak_ke_C1, Jarak_ke_C2, Jarak_ke_C3,
         Alasan)

# Print hasil
penjelasan
```

5. Perhitungan Pembagian Jenis *Cluster*

```
{r}
# Data numerik untuk perhitungan cluster
num_data <- data %>%
  select(Fisik, Psikis, Seksual, Eksploitasi, TPPO, Penelantaran, Lainnya)

# Jalankan ulang kmeans
set.seed(123)
km <- kmeans(num_data, centers = 3)

# Tambahkan cluster ke data
data$Cluster <- km$cluster

# Hitung rata-rata setiap variabel per cluster
cluster_mean <- aggregate(num_data, list(Cluster = km$cluster), mean)

# Hitung total nilai tiap cluster
cluster_mean$Total_Skor <- rowSums(cluster_mean[,2:8])

# Ranking cluster berdasarkan total skor
ranking <- order(cluster_mean$Total_Skor)

# Tambahkan label kategori
cluster_mean$Kategori <- NA
cluster_mean$Kategori[ranking[1]] <- "Kasus Rendah"
cluster_mean$Kategori[ranking[2]] <- "Kasus Sedang"
cluster_mean$Kategori[ranking[3]] <- "Kasus Tinggi"

# Tampilkan hasil perhitungan
cluster_mean
```

6. Visualisasi PCA Menggunakan Map Sumatera Selatan

```
{r}
library(sf)
library(dplyr)
library(ggplot2)
library(stringr)
library(lwgeom)

# 1. LOAD SHAPEFILE
sumsel <- st_read("C:/Users/muham/Downloads/RBI_50K_2023_Sumatera Selatan
.shp")

# 2. FIX GEOMETRY (WAJIB)
sumsel <- st_make_valid(sumsel)

# 3. STANDARISASI NAMA
sumsel$Cakupan <- sumsel$WADMKK %>%
  toupper() %>%
  str_replace("KABUPATEN ", "") %>%
  str_replace("KOTA ", "") %>%
  str_trim()

# 4. DATA CLUSTER
data_cluster <- obj$data %>%
  mutate(
    Cakupan = toupper(Cakupan) %>%
      str_replace("KABUPATEN ", "") %>%
      str_replace("KOTA ", "") %>%
      str_trim()
  )

# 5. MERGE
sumsel_cluster <- sumsel %>%
  left_join(data_cluster, by = "Cakupan")

# 6. CENTROID
sumsel_cluster$centroid <- st_centroid(sumsel_cluster$geometry)
sumsel_cluster$label <- sumsel_cluster$Cakupan
```

```

# 7. PETA
ggplot(sumsel_cluster) +
  geom_sf(aes(fill = Cluster_Label), color = "white", size = 0.3) +
  geom_sf_text(aes(label = label, geometry = centroid),
               size = 2, color = "black", check_overlap = FALSE) +
  scale_fill_manual(
    values = c("Kasus Rendah" = "#9FE2BF",
               "Kasus Sedang" = "#FFD580",
               "Kasus Tinggi" = "#FF6961")
  ) +
  labs(
    title = "Peta Klaster Kekerasan Terhadap Anak di Sumatera Selatan",
    fill = "Kategori Kasus"
  ) +
  theme_minimal()

```