



PENERAPAN METODE CLUSTER DALAM DATA MINING MENGELOMPOKKAN KENAKALAN REMAJA (STUDI KASUS POLDA SULTRA)

Muhammad Rival¹, Misriani², La Ode Bakrim³

¹muhammadrabiul828@gmail.com, ²misrianiruufa87@gmail.com, ³vbakrim@gmail.com
^{1,2,3}STIMIK Bina Bangsa Kendari

Abstrak

Masalah sosial perilaku menyimpang kenakalan remaja di wilayah Sulawesi Tenggara sudah banyak terdengar atau bahkan sering membuat masyarakat itu sendiri merasakan dampak yang terjadi akibat kenakalan remaja tersebut. Data Mining yaitu sebuah proses menyatukan dan mengolah data untuk mengekstrak informasi penting. K-Means clustering adalah algoritma *unsupervised learning* yang dipakai untuk mengelompokkan *dataset* yang belum dilabel ke dalam kluster yang berbeda. Simbol K pada K-Means *clustering* menandakan jumlah kluster yang digunakan. Hasil dari penelitian ini dengan wilayah yang terdominasi jumlah kenakalan remaja terbanyak yaitu kota Kendari yang menduduki puncak sebanyak 2588 kasus, Muna 2427 kasus, Baubau 893 kasus, Konawe 786 kasus, Konawe Selatan 776 kasus, Buton 558 kasus, Kolaka 494 kasus, Wakatobi 240 kasus, Kolaka Utara 212 kasus, Bombana 187 kasus, Butur 90 kasus, Konawe Utara 82 kasus, Buton Tengah 14 kasus, dan Kolaka Timur sebanyak 12 kasus kenakalan remaja.

Kata kunci: K-Means Clustering, Data Mining, Pengelompokan, Kenakalan Remaja

Abstract

The social problem of deviant behavior by juvenile delinquents in the Southeast Sulawesi region has been widely heard about or often even makes the community itself feel the impact that occurs due to juvenile delinquency. Data Mining is a process of combining and processing data to extract important information. K-Means clustering is an unsupervised learning algorithm that is used to group unlabeled datasets into different clusters. The K symbol in K-Means clustering indicates the number of clusters used. The results of this research are the areas dominated by the highest number of juvenile delinquents, namely the city of Kendari which topped with 2588 cases, Muna 2427 cases, Baubau 893 cases, Konawe 786 cases, South Konawe 776 cases, Buton 558 cases, Kolaka 494 cases, Wakatobi 240 cases, North Kolaka 212 cases, Bombana 187 cases, Butur 90 cases, North Konawe 82 cases, Central Buton 14 cases, and East Kolaka 12 cases of juvenile delinquency.

Keywords: *K-means Clustering, Data Mining, Grouping, Juvenile Delinquency*

1. Pendahuluan

Masalah sosial perilaku menyimpang kenakalan remaja di wilayah Sulawesi Tenggara sudah banyak terdengar atau bahkan sering membuat masyarakat itu sendiri merasakan dampak yang terjadi akibat kenakalan remaja tersebut. Kenakalan remaja yang sering terjadi antara lain pencurian, perkelahian, penganiayaan, pengeroyokan dan penyalahgunaan obat-obatan terlarang. Hal tersebut berdasarkan laporan dari masyarakat atau bahkan dari pihak sekolah yang mendapat kasus terhadap siswanya. Dalam kasus tersebut banyak pihak yang merasa dirugikan akibat kenakalan yang dilakukan para pelaku kenakalan tersebut, mulai dari sekolah itu sendiri, para orang tua, atau bahkan para korban dari perilaku kenakalan remaja. Sehingga tak jarang timbul rasa dendam dari teman sebaya akibat kenakalan yang dilakukan oleh kelompok lain dan menimbulkan masalah berlanjut hingga melakukan tawuran antar kelompok.

Penelitian ini dimaksudkan untuk mengetahui lebih dalam mengenai kenakalan remaja di wilayah Sulawesi Tenggara. Melalui studi kasus ini juga peneliti ingin mengetahui kelompok-kelompok remaja yang sering terlibat aktif dalam kasus kenakalan remaja mulai dari karakteristik dan perbedaan usia dari antar kelompok kenakalan yang terjadi. Kenakalan remaja dapat memiliki dampak negatif yang signifikan pada masyarakat dan keamanan di wilayah Sulawesi Tenggara. Hal ini dapat mencakup peningkatan kejahatan, ketidakamanan masyarakat, dan gangguan ketertiban. Dengan melakukan analisis *clustering*, penelitian ini dapat membantu memahami pola-pola karakter dan perbedaan kenakalan remaja yang ada dan mengidentifikasi faktor-faktor yang berkontribusi terhadapnya.

Dari latar belakang tersebut penulis menyimpulkan bahwa kenakalan yang terjadi akibat kurangnya perhatian dari para orang tua atau pihak yang memiliki pengaruh untuk mendidik dan membuat kenakalan remaja yang terjadi menjadi diwilayah tersebut menjadi terkendali. Dengan ini penulis mengambil judul yaitu “Penerapan Metode Cluster Dalam Data Mining Mengelompokkan Kenakalan Remaja (Studi Kasus Polda Sultra)”. Dalam sebuah penelitian dimana yang menjadi objek penelitian yaitu kepada pihak yang sering menerima kasus atau pihak yang sebagai pengamanan di wilayah tersebut yaitu Polda Sulawesi Tenggara.

Dengan menggunakan metode *clustering*, sehingga dapat mengidentifikasi karakteristik kelompok data dan perbedaan kenakalan pada remaja yang ada di daerah tersebut. Hal ini dapat membantu mengidentifikasi objek-objek berdasarkan karakteristik dan untuk mendukung sistem pendukung keputusan. Untuk menentukan kelompok data kenakalan remaja yang teridentifikasi berdasarkan analisis *clustering* sesuai dengan jumlah *cluster* yang ditentukan. Tujuan tersebut yaitu untuk memperoleh pengetahuan yang baik tentang karakteristik dan perbedaan antara kelompok-kelompok kenakalan remaja yang ada dengan berdasarkan analisis *clustering*.

Dari Pramudiono [1] Data Mining yaitu sebuah proses buat mengetahui sesuatu yang bernilai pada kumpulan data semisal pengetahuan yang pada saat ini belum ditemukan baik itu manual atau secara otomatis asal data yang diterima dan kompleks dengan maksud untuk menentukan pola-pola atau kebiasaan yang krusial yang sering kurang disadari keberadaannya. Atau sebuah proses untuk mengetahui nilai tambahan berupa ilmu yang bermanfaat yang pada ini kurang ditemui melalui manual berasal dari gugusan sebuah data. Sebuah data mining juga kadang dianggap (*Knowledge Discovery in Database*). Dimana artinya sebuah usaha menutupi sekumpulan data, pemakaian, dan historis untuk menentukan ketentuan dasar atau sejalan pada data set yang besar.[2]

2. Metode

K-means *clustering* adalah sebuah metode dalam analisis data yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan karakteristik mereka. Tujuan utama dari K-means *clustering* adalah untuk membagi data ke dalam kelompok-kelompok sedemikian rupa sehingga data dalam satu kelompok memiliki kesamaan antara satu sama lain, sementara data dari kelompok yang berbeda memiliki perbedaan yang signifikan [3].

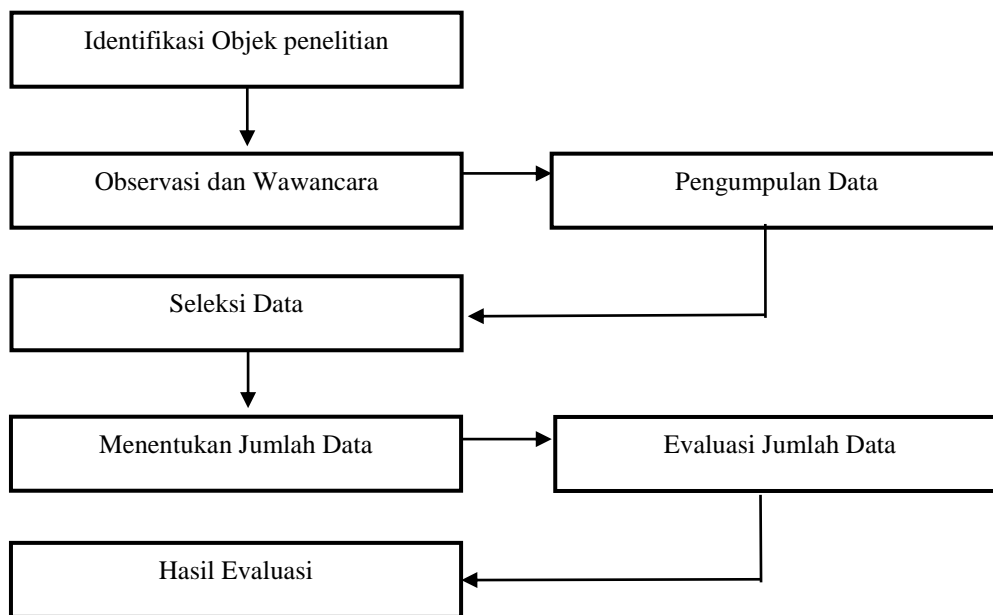
Hasil akhir dari K-means *clustering* adalah sejumlah K kelompok dengan data yang memiliki kesamaan karakteristik dalam setiap kelompok tersebut[4]. Metode ini digunakan dalam berbagai aplikasi seperti analisis data, segmentasi pelanggan, pengelompokan berita, pengolahan citra, dan banyak lagi.

Tabel 1. Penelitian Terdahulu

Nama Peneliti	Judul	Metode	Hasil
Muhammad Zulfadhilah [5]	Implementasi Metode K Means Clustering untuk Meningkatkan Penjangkaran Mahasiswa	Clustering K-means Rapidminer dan orange data mining	Mengelompokkan mahasiswa berdasarkan asal sekolah kota dan provinsi dengan menggunakan aplikasi rapidminer dan orange data mining
Tifani Amalina [6]	Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food	Algoritma Clustering K-means	Penerapan algoritma K-Means Clustering pada data penjualan produk Frozen Food, untuk menghasilkan sebuah informasi mengenai data pengelompokan minat konsumen tertinggi dan terendah.

Tb Munandar [7]	Ai Penerapan algoritma clustering untuk pengelompokan tingkat kemiskinan provinsi banten	Algoritma Clustering K-means	Mengelompokkan Suatu daerah berdasarkan tingkat kemiskinan mulai dari tingkat yang tertinggi hingga ketinggian yang paling rendah
Asmana [8]	Clustering data calon siswa baru menggunakan metode k-means di sekolah menengah kejuruan wahidin kota cirebon	Clustering K-means Rapidminer	Mengalalisis clustering data asal calon siswa mencari informasi tentang penerimaan peserta didik baru di SMK Wahidin Kota Cirebon .
Tajunisha [9]	Analisis kinerja k-means dengan metode inialisasi yang berbeda untuk data dimensi tinggi	Principal Component Analisis (PCA) dan Algoritma K-means	Hasil akhir yang didapatkan, terlihat bahwa metode baru yang diusulkan lebih baik daripada metode yang sebelumnya.

Metodologi penelitian ini dilakukan agar memperoleh sesuatu secara sistematis dengan menggunakan cara atau metode ilmiah. Metodologi penelitian ini merupakan gambaran sebuah langkah-langkah yang spesifik untuk dilakukan selama dalam proses penelitian. Metodologi penelitian yang digunakan, dapat berupa alur berikut ini:

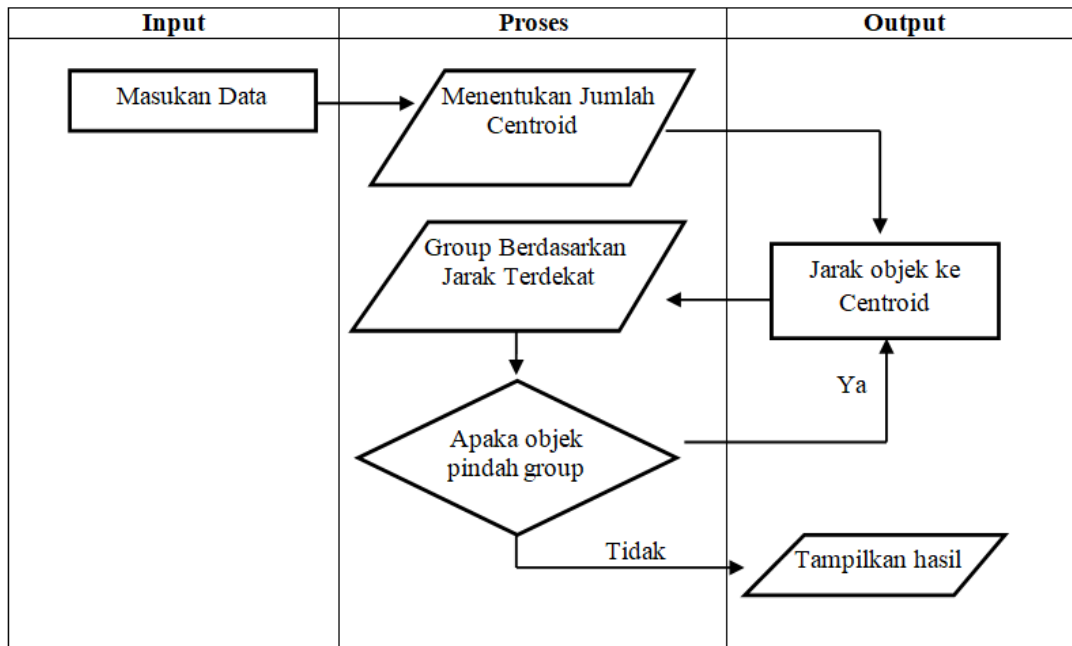


Gambar 1. Alur Kerja Penelitian

Objek penelitian adalah suatu topik atau sasaran ilmiah yang akan diteliti untuk mendapatkan informasi dan data untuk tujuan penelitian. Observasi atau Wawancara yaitu melakukan pengamatan atau melibatkan interaksi secara langsung sehingga dapat memberikan berupa informasi persepsi dan pengalaman subjek. Pengumpulan data berlaku saat setelah hasil observasi dan wawancara pada objek penelitian. Seleksi data dalam data yang telah dikumpulkan bertujuan untuk memperoleh data yang akurat dan bertujuan untuk mengambil data tertentu untuk diproses sesuai data yang diinginkan. Menentukan jumlah data ini bertujuan untuk memilih data yang dominan dari semua jumlah data yang ada untuk tingkat kenakalan. Setelah jumlah data telah ditentukan kemudian data dilihat apakah telah sesuai berdasarkan hasil seleksi sehingga dapat diprogram dengan baik oleh system.

3. Hasil dan Pembahasan

Pada sistem yang terbentuk merujuk pada proses evaluasi dan pemahaman terhadap sistem yang sedang beroperasi secara aktif. Sistem dalam konteks ini bisa merujuk pada sistem komputer, sistem bisnis, atau sistem lainnya yang melibatkan berbagai komponen, proses, dan interaksi.



Gambar 2. Alur Sistem Yang Berjalan

Pada tahap untuk memulai program dengan menginput data atribut yang akan dikelompokkan. Misal atribut data usia, jenis kelamin, dan jenis kenakalan. Menentukan jumlah *centroid* berdasarkan inputan data. Misalnya, terdapat dua pasien dan yang menjadi *centroidnya* yaitu wilayah, jenis kenakalan, dan usia kemudian ditransformasikan. Setelah ditransformasi kemudian menentukan *centroid* awal berdasarkan dari data pasien 1 dan 2. Penentuan *centroid* dapat diambil secara acak.

Kemudian hitung pada jarak sebuah data dengan C-1 atau *centroid* sesuai jumlah atribut dengan *Euclidean Distance*. Group berdasarkan jarak terdekat: Mengelompokkan data sesuai pada jarak terdekat atau yang paling rendah (*Distance Minimum*). Apakah objek pindah? Merumuskan sebuah perulangan dimana kondisi sehingga apakah posisi *centroid* tetap dan tidak ada perubahan terhadap datanya? Jika Ya, maka telah menyelesaikan pengelompokan. Jika masih ada perubahan *centroid* maka update kembali nilai *centroid* melalui proses pertama. Jika sudah tidak ada perubahan lagi maka menampilkan pengelompokan pada data dan *centroid* berdasarkan *cluster* masing-masing.

Dalam sebuah konteks penelitian, teknik wawancara berguna bagi peneliti untuk mendapatkan informasi mendalam tentang sebuah topik tertentu, juga pemahaman yang bersifat subjektif, pandangan, pengalaman, atau sebuah pandangan antar individu dan/ atau kelompok. Sehingga dalam penerapannya peneliti dapat memperoleh data yang valid dari pihak yang berwenang salah satunya adalah data Kenakalan Remaja.

Tabel 1. Data Kenakalan Remaja

TH	JK	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
2018	A	3	142	34	39	19	13	93	13	0	0	3	0	0	0	0
2018	B	26	303	60	101	94	38	275	64	29	24	6	0	0	0	0
2018	C	44	185	6	73	38	21	84	13	28	18	17	0	0	0	0
2019	A	11	118	40	26	12	32	115	21	2	0	7	0	0	0	0
2019	B	25	221	61	109	91	92	288	70	21	13	13	0	0	0	0
2019	C	25	101	16	46	49	34	80	24	5	17	8	0	0	0	0
2020	A	3	83	20	26	26	32	96	16	0	4	6	4	9	0	0
2020	B	10	196	50	55	99	91	276	96	18	20	19	8	13	0	0
2020	C	22	83	12	30	53	45	57	14	13	6	10	2	5	0	0
2021	A	3	122	11	19	21	29	85	9	1	0	3	5	13	0	0
2021	B	10	210	29	62	84	80	247	62	13	30	21	38	27	0	0
2021	C	25	37	12	41	30	35	79	21	6	2	8	8	7	0	0
2022	A	2	116	16	23	15	43	79	12	0	8	5	1	2	0	0
2022	B	18	214	80	60	106	103	293	84	29	59	33	5	7	0	0
2022	C	26	23	10	37	64	14	38	9	5	8	13	1	0	0	0
2023	A	0	34	12	10	14	13	57	12	1	5	1	8	1	2	2
2023	B	15	105	24	22	42	53	155	42	10	22	22	1	5	10	9
2023	C	13	14	1	7	36	8	30	3	6	4	17	1	1	0	3

Keterangan :

- | | | | |
|--------------|--------------------|----------------------|---------------------|
| X1 = Polda | X9 = Bombana | JK = Jenis Kejahatan | C1 = Cluster1 |
| X2 = Kendari | X10 = Wakatobi | TH = Tahun | C2 = Cluster2 |
| X3 = Kolaka | X11 = Kolaka utara | A = Pengeroyokan | JT = Jarak Terdekat |
| X4 = Konawe | X12 = Konawe utara | B = Aniaya biasa | K = Kelompok |
| X5 = Bau bau | X13 = Buton utara | C = Curi biasa | |
| X6 = Konsel | X14 = Kolaka timur | | |
| X7 = Muna | X15 = Buton tengah | | |
| X8 = Buton | | | |

Berdasarkan transformasi kriteria jenis kenakalan remaja, wilayah dan tahun terjadinya kenakalan remaja, sehingga dapat mempermudah peneliti mengolah data dalam mencari nilai cluster data, menentukan nilai titik *cluster*, hitung jarak data pada titik *cluster (Centroid)* dan mencari kelompok data. Dari transformasi tersebut, kemudian menentukan pusat atau titik awal pada *Cluster (Centroid)*. Dalam menentukan titik awal diasumsikan:

- Data diperoleh dari data kedua untuk pusat data pada c-1 yaitu: (26, 303, 60, 181, 94, 38, 275, 64, 29, 24, 6, 0, 0, 0)
- Data ke 3 diambil sebagai pusat data cluster 2 yaitu: (44, 185, 6, 73, 38, 21, 84, 13, 28, 18, 17, 0, 0, 0)

Pusat data cluster 1 dan 2 diambil secara acak dalam penentuan centroid.

Iterasi Pertama

Tabel 2. Pusat Data Cluster

C1	26	303	60	101	94	38	275	64	29	24	6	0	0	0	0
C2	44	185	6	73	38	21	84	13	28	18	17	0	0	0	0

Pada metode *Clustering* setiap *cluster* memiliki titik pusat *cluster (Centroid)* dan secara umum metode ini memiliki fungsi yaitu meminimumkan jarak (*dissimilarity*) dari seluruh data ke pusat *cluster* masing-masing.

Tabel 3. Klusterisasi Data Iterasi Pertama

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	C1	C2	JT	K
3	142	34	39	19	13	93	13	0	0	3	0	0	0	0	73865	5640	5640	C2
26	303	60	101	94	38	275	64	29	24	6	0	0	0	0	0	60307	0	C1
44	185	6	73	38	21	84	13	28	18	17	0	0	0	0	60307	0	C2	
11	118	40	26	12	32	115	21	2	0	7	0	0	0	0	75780	10809	10809	C2
25	221	61	109	91	92	288	70	21	13	13	0	0	0	0	10154	58441	10154	C1
25	101	16	46	49	34	80	24	5	17	8	0	0	0	0	88061	8942	8942	C2
3	83	20	26	26	32	96	16	0	4	6	4	9	0	0	95991	14466	14466	C2
10	196	50	55	99	91	276	96	18	20	19	8	13	0	0	18079	55130	18079	C1
22	83	12	30	53	45	57	14	13	6	10	2	5	0	0	108128	14289	14289	C2
3	122	11	19	21	29	85	9	1	0	3	5	13	0	0	88007	8764	8764	C2
10	210	29	62	84	80	247	62	13	30	21	38	27	0	0	16489	38434	16489	C1
25	37	12	41	30	35	79	21	6	2	8	8	7	0	0	122161	24266	24266	C2
2	116	16	23	15	43	79	12	0	8	5	1	2	0	0	91502	9475	9475	C2
18	214	80	60	106	103	293	84	29	59	33	5	7	0	0	17131	68594	17131	C1
26	23	10	37	64	14	38	9	5	8	13	1	0	0	0	146548	31077	31077	C2
0	34	12	10	14	13	57	12	1	5	1	8	1	2	2	141468	29447	29447	C2
15	105	24	22	42	53	155	42	10	22	22	1	5	10	9	65393	16848	16848	C2
13	14	1	7	36	8	30	3	6	4	17	1	1	0	3	164922	37533	37533	C2

Dalam klusterisasi data setelah memilih *Centroid* dimana itu adalah pusat dari *cluster*. Selanjutnya menghitung jarak data dengan rumus *Euclidean* yaitu sebagai berikut:

JT = Jarak Terdekat

K = Kelompok

$$D(i,j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ik} - X_{jk})^2} [10]$$

Dimana :

$D(i,j)$ = Jarak data ke i ke pusat *cluster* j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

Sehingga memperoleh:

$$C1 = D(i,j)$$

$$D(i,j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ik} - X_{jk})^2}$$

$$D(i,j) = \text{SQRT}((D19-\$D\$23)^2) + ((E19-\$E\$23)^2) + ((F19-\$F\$23)^2) + ((G19-\$G\$23)^2) + ((H19-\$H\$23)^2) + ((I19-\$I\$23)^2) + ((J19-\$J\$23)^2) + ((K19-\$K\$23)^2) + ((L19-\$L\$23)^2) + ((M19-\$M\$23)^2) + ((N19-\$N\$23)^2) + ((O19-\$O\$23)^2) + ((P19-\$P\$23)^2) + ((Q19-\$Q\$23)^2) + ((R19-\$R\$23)^2).$$

C1 = Data ke ((2, 5, 8, 11, 14)

$$C2 = D(i,j)$$

$$D(i,j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ik} - X_{jk})^2}$$

$$D(i,j) = \text{SQRT}((D2-\$D\$24)^2) + ((E2-\$E\$24)^2) + ((F2-\$F\$24)^2) + ((G2-\$G\$24)^2) + ((H2-\$H\$24)^2) + ((I2-\$I\$24)^2) + ((J2-\$J\$24)^2) + ((K2-\$K\$24)^2) + ((L2-\$L\$24)^2) + ((M2-\$M\$24)^2) + ((N2-\$N\$24)^2) + ((O2-\$O\$24)^2) + ((P2-\$P\$24)^2) + ((Q2-\$Q\$24)^2) + ((R2-\$R\$24)^2)$$

C2 = Data ke (1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 17, 18).

Pada iterasi pertama dapat disimpulkan bahwa untuk kelompok *Centroid* pertama dan kedua terdapat 5 data (2, 5, 8, 11, 14) dan kelompok *Centroid* 2 terdapat 13 data (1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 17, 18).

Setelah iterasi pertama dapat diketahui kelompok *Centroid*-nya sehingga untuk mengetahui apakah objek berpindah group atau tidak maka dapat membuat iterasi ke 2. Jika objek berpindah maka akan digroup dengan jarak yang paling dekat lagi. Apabila tetap atau tidak mengalami perubahan maka iterasi selesai.

Iterasi Kedua

Tabel 4. Pusat Cluster Baru

C1	17,8	228,8	56	77,4	94,8	80,8	275,8	75,2	22	29,2	18,4	10,2	9,4	0	0
C2	14,769	89,461	16,461	30,692	32,230	28,615	80,615	16,077	5,923	7,230	9,230	2,384	3,307	0,923	1,07692

Pusat data cluster baru dapat diketahui dengan menjumlahkan data yang ada di Cluster 1 yang sebelumnya telah diketahui diiterasi pertama yaitu untuk Centroid 1 data ke (2, 5, 8, 11, 14) dan Centroid 2 yaitu data ke (1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 17, 18). Sehingga setelah dijumlah kemudian dibagi dari masing-masing jumlah data, untuk $C1 = (2, 5, 8, 11, 14)/5$ dan $C2 = (1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 17, 18)/13$. Sehingga dari total penjumlahan yang dilakukan dapat diketahui pusat data *cluster* baru.

Tabel 5. Clasterisasi Data Iterasi Kedua

TH	JK	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	C1	C2	JT	K
2018	A	3	142	34	39	19	13	93	13	0	0	3	0	0	0	0	243,1	63,3	63,26	C2
2018	B	26	303	60	101	94	38	275	64	29	24	6	0	0	0	0	92,34	312	92,34	C1
2018	C	44	185	6	73	38	21	84	13	28	18	17	0	0	0	0	230,1	113	112,6	C2
2019	A	11	118	40	26	12	32	115	21	2	0	7	0	0	0	0	234,2	55,9	55,89	C2
2019	B	25	221	61	109	91	92	288	70	21	13	13	0	0	0	0	44,01	281	44,01	C1
2019	C	25	101	16	46	49	34	80	24	5	17	8	0	0	0	0	254,9	31	31,02	C2
2020	A	3	83	20	26	26	32	96	16	0	4	6	4	9	0	0	264	24,4	24,39	C2
2020	B	10	196	50	55	99	91	276	96	18	20	19	8	13	0	0	48,45	258	48,45	C1
2020	C	22	83	12	30	53	45	57	14	13	6	10	2	5	0	0	284,4	37,9	37,86	C2
2021	A	3	122	11	19	21	29	85	9	1	0	3	5	13	0	0	259,9	42,2	42,16	C2
2021	B	10	210	29	62	84	80	247	62	13	30	21	38	27	0	0	60,6	231	60,6	C1
2021	C	25	37	12	41	30	35	79	21	6	2	8	8	7	0	0	298,6	56	55,96	C2
2022	A	2	116	16	23	15	43	79	12	0	8	5	1	2	0	0	263,3	38,8	38,85	C2
2022	B	18	214	80	60	##	103	293	84	29	59	33	5	7	0	0	57,2	291	57,2	C1
2022	C	26	23	10	37	64	14	38	9	5	8	13	1	0	0	0	336,8	88	88	C2
2023	A	0	34	12	10	14	13	57	12	1	5	1	8	1	2	2	330,4	70,9	70,91	C2
2023	B	15	105	24	22	42	53	155	42	10	22	22	1	5	10	9	197,9	88,4	88,41	C2
2023	C	13	14	1	7	36	8	30	3	6	4	17	1	1	0	3	360	98,8	98,76	C2

Pada iterasi kedua untuk menentukan C1 dan C2 kemudian jarak terdekat Cluster dan juga kelompok cluster, maka semua data awal dihitung ke Cluster yang baru sehingga dapat memperoleh nilai untuk kelompok C1 dan C2. Sehingga pada iterasi kedua ini dapat diketahui apakah objek pindah group atau tidak. Jika berpindah group maka dilakukan menghitung jarak objek ke Centroid lagi dan jika iterasi kedua tidak berpindah group maka Klusterisasi selesai.

Dalam menentukan iterasi kedua sama seperti yang dilakukan pada rumus iterasi pertama yaitu dengan menjumlahkan data awal dengan pusat Cluster. Rumus yang digunakan yaitu Algoritma *Euclidean Distance*.

$$D(i,j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Sehingga dapat diketahui:

$$C1 = D(i,j)$$

$$D(i,j) = \text{SQRT}(((D32-\$D\$27)^2) + ((E32-\$E\$27)^2) + ((F32-\$F\$27)^2) + ((G32-\$G\$27)^2) + ((H32-\$H\$27)^2) + ((I32-\$I\$27)^2) + ((J32-\$J\$27)^2) + ((K32-\$K\$27)^2) + ((L32-\$L\$27)^2) + ((M32-\$M\$27)^2) + ((N32-\$N\$27)^2) + ((O32-\$O\$27)^2) + ((P32-\$P\$27)^2) + ((Q32-\$Q\$27)^2) + ((R32-\$R\$27)^2))$$

C1 = Data ke ((2, 5, 8, 11, 14)

C2 Dapat diketahui:

$$C2 = \text{SQRT}(((D32-\$D\$28)^2) + ((E32-\$E\$28)^2) + ((F32-\$F\$28)^2) + ((G32-\$G\$28)^2) + ((H32-\$H\$28)^2) + ((I32-\$I\$28)^2) + ((J32-\$J\$28)^2) + ((K32-\$K\$28)^2) + ((L32-\$L\$28)^2) + ((M32-\$M\$28)^2) + ((N32-\$N\$28)^2) + ((O32-\$O\$28)^2) + ((P32-\$P\$28)^2) + ((Q32-\$Q\$28)^2) + ((R32-\$R\$28)^2))$$

C2 = Data ke (1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 17, 18)

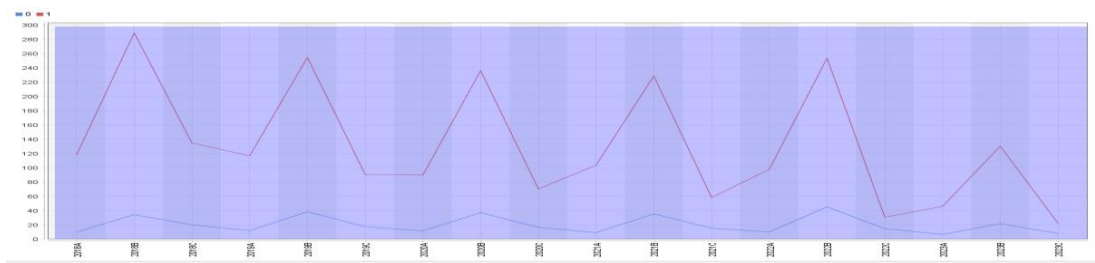
Dari iterasi yang pertama dan kedua objek data tidak mengalami perpindahan group yang artinya data yang dihasilkan dari pengelompokan Klusterisasi telah memperoleh nilai dan data pengelompokan yang sesuai. Maka dari itu, Klusterisasi pada data awal dapat memperoleh kelompk data berdasarkan jarak terdekat.

Adapun hasil pada Klusterisasi terdapat dua cluster yaitu cluster 0 dan cluster 1 dimana masing-masing sama-sama memiliki atribut. Pada cluster 0 terdapat 5 atribut yaitu data ke 2, 5, 8, 11, 14. Untuk cluster 1 terdapat 13 atribut yaitu data ke 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 17, 18.

Tabel 6. Tampilan Hasil Transkripsi

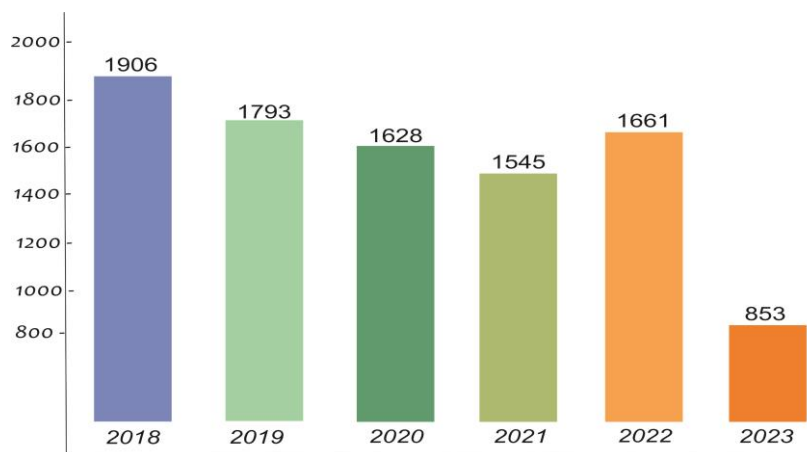
cluster 1	3	143	34	39	19	13	93	13	0	0	3	0	0	0
cluster 0	26	303	60	101	94	38	75	64	29	24	6	0	0	0
cluster 1	44	185	6	73	38	21	84	13	28	18	17	0	0	0
cluster 1	11	118	40	26	12	32	115	21	2	0	7	0	0	0
cluster 0	25	221	61	109	91	92	228	70	21	13	13	0	0	0
cluster 1	25	101	16	46	49	34	80	24	5	17	8	0	0	0
cluster 1	3	83	20	26	26	32	96	16	0	4	6	4	9	0
cluster 0	10	196	50	55	99	91	276	96	18	20	19	8	13	0
cluster 1	22	83	12	30	53	45	57	14	13	6	10	2	5	0
cluster 1	3	122	11	19	21	29	85	9	1	0	3	5	13	0
cluster 0	10	210	29	62	84	80	247	62	13	30	21	38	27	0
cluster 1	25	37	12	41	30	35	79	21	6	2	8	8	7	0
cluster 1	2	116	16	23	15	43	79	12	0	8	5	1	2	0
cluster 0	18	214	80	60	106	103	293	84	29	59	33	5	7	0
cluster 1	26	23	10	37	64	14	38	9	5	8	13	1	0	0
cluster 1	0	34	12	10	14	13	57	12	1	5	1	8	1	2
cluster 1	15	105	24	22	42	53	155	42	10	22	22	1	5	10
cluster 1	13	14	1	7	36	8	30	3	6	4	17	1	1	0

Pada kurva grafik yang ditampilkan, pada Tahun 2018 memiliki data rata-rata yang terbesar dari jumlah data yang ada. Dan kurva yang paling terkecil terdapat pada kurva tahun 2023 karena berdasarkan data yang didapat minimnya perolehan jumlah atribut kenakalan yang masuk.

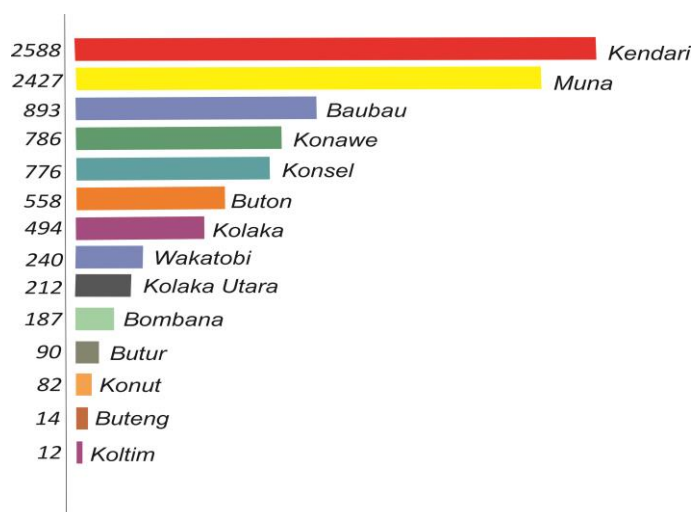


Gambar 4. Tingkatan Kenakalan Remaja

Dari beberapa tahun terakhir nampaknya dari tahun sebelumnya hingga tahun berikutnya kenakalan remaja yang terjadi mengalami penurunan secara bertahap. Hingga ditahun 2023 total kenakalan remaja yang terjadi hanya berkisar 853 saja dari tahun sebelumnya yang memiliki predikat terbanyak ditahun 2018 yaitu 1906 kasus.



Gambar 5. Jumlah Kenakalan Remaja 6 Tahun Terakhir di Wilayah Sulawesi Tenggara



Gambar 6. Grafik Wilayah Jumlah Kanakalan Remaja Tertinggi Periode 2018 hingga 2023

Pada grafik yang ditampilkan menunjukkan jumlah kasus kenakalan remaja dari 6 tahun terakhir dengan memperlihatkan jumlah kasus tertinggi dan paling minim diberbagai wilayah Sulawesi tenggara. Puncak kenakalan remaja yang berada dikota kendari mencapai 2588 kasus. Dengan begitu fokus penanganan dan pengamanan kenakalan remaja dapat di tingkatkan berdasarkan keluaran jumlah tertinggi pada wilayah yang potensi terjadinya kenakalan remaja tersebut.

Dari percobaan Clustering diatas hasil yang didapatkan adalah kenakalan remaja yang terjadi ditahun 2018 menjadi puncak terbanyaknya kasus penganiayaan dengan angka terbanyak 1906. Ditahun 2019 memperoleh 1793 kasus, tahun 2020 memperoleh 1628 kasus, tahun 2021 memperoleh 1545 kasus, tahun 2022 memperoleh 1661 kasus, dan ditahun 2023 memperoleh 853 kasus kenakalan remaja.

Dengan demikian wilayah yang terdominasi jumlah kenakalan remaja terbanyak yaitu Kota Kendari yang menduduki puncak sebanyak 2588 kasus, kedua Muna 2427 kasus, ketiga Baubau 893 kasus, Konawe 786 kasus, keempat Konawe Selatan 776 kasus, kelima Buton 558 kasus, keenam Kolaka 494 kasus, ketujuh Wakatobi 240 kasus, kedelapan Kolaka Utara 212 kasus, kesembilan Bombana 187 kasus, kesepuluh Buton Utara 90 kasus, kesebelas Konawe Utara 82 kasus, kedua belas Buton Tengah 14 kasus, dan terakhir Kolaka Timur Sebanyak 12 kasus kenakalan remaja.

4. Kesimpulan dan Saran

4.1. Kesimpulan

Berdasarkan penelitian yang dilakukan, pengelompokan data yang terbentuk antara kelompok-kelompok kenakalan remaja sesuai dengan analisis *clustering* dengan jarak data ke *pusat* cluster yaitu:

1. Kasus penganiayaan menjadi kasus kenakalan remaja yang paling banyak terjadi dari periode tahun ke tahun dengan memperoleh 5 data kasus terbesar berdasarkan analisis *clustering*.
2. Kota Kendari menjadi wilayah yang paling dominan mendapatkan kasus kenakalan remaja dan Kolaka Timur menjadi daerah yang memiliki tingkat kenakalan paling minim.
3. Kasus kenakalan remaja yang terjadi mengalami penurunan jumlah kasus dari tahun ke tahun. Tercatat ditahun 2023 jumlah kasus secara keseluruhan hanya memperoleh 853 jenis kenakalan remaja berbeda ditahun sebelumnya yang menembus angka seribu kasus kenakalan remaja.

4.2. Saran

Penelitian yang dilakukan oleh peneliti hanya pada data kenakalan yang terkategori saja untuk jumlah kenakalan yang dominan terjadi di wilayah Sulawesi tenggara. Untuk dapat memperoleh hasil maksimal diharapkan peneliti selanjutnya dapat menggunakan metode lain seperti *Hierarchical Clustering (Connectivity Based Cluster)*, *Density-based Clustering*, *Distribution-based Clustering*. Sehingga dapat memperoleh dan mengolah data lebih banyak lagi melebihi daripada jumlah data yang didapatkan dalam penelitian ini.

5. Ucapan Terima Kasih

Peneliti banyak berterimakasih kepada pihak dawan reskrim polda sultra yang telah banyak memberikan banyak informasi, saran dan masukkan untuk proses jalannya penelitian ini. Dengan informasi yang didapatkan sehingga dapat mampu membuat banyak argumen ilmiah untuk pembuatan karya ilmiah pada penelitian ini.

Daftar Pustaka

- [1] H. R. M. Tambun and A. Anofrizen, (vol. 1, no. 1, pp. 31–38, 2015) “Rancang Bangun Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, [Online]. Available: http://ejournal.uin-suska.ac.id/index.php/RMSI/article/view/1302/pdf_3
- [2] S. Kasus and P. Binjai, (vol.1, no.1 August, 2018), “DATA MINING PENGELOMPOKAN KASUS KENAKALAN REMAJA DI KOTA BINJAI DENGAN MENGGUNAKAN METODE CLUSTERING (STUDI KASUS POLRES BINJAI) Latar Belakang Masalah kenakalan remaja pada saat ini semakin meresahkan masyarakat , kenakalan yang dilakukan oleh para pem,”
- [3] T. Rismawan and D. S. Kusumadewi, (vol. 21, no. 01, pp. 1907–5022, 2008.)“Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (Bmi) & Ukuran Kerangka,” *Semin. Nas. Apl. Teknol. Inf.*

- [4] T. Steinbach, (2004). “Data Mining Cluster Analysis : Basic Concepts and Algorithms - Lecture Notes,”
- [5] Muhammad Zulfadhilah, Mambang, and Septyan Eka Prastya, (vol. 9, no. 2, pp. 152–160, 2022). “Implementasi Metode K-Means Clustering untuk Meningkatkan Penjarangan Mahasiswa,” *Tematik*, doi: 10.38204/tematik.v9i2.1053.
- [6] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, (vol. 8, no. 15, pp. 574–583, 2022). “Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food,” *J. Ilm. Wahana Pendidik.*, [Online]. Available: <https://doi.org/10.5281/zenodo.7052276>
- [7] T. A. Munandar, (vol. 9, no. 2, pp. 109–114, 2022). “Penerapan Algoritma Clustering Untuk Pengelompokan Tingkat Kemiskinan Provinsi Banten,” *JSiI (Jurnal Sist. Informasi)*, doi: 10.30656/jsii.v9i2.5099.
- [8] Y. A. Wijaya *et al.*, (vol. 6, no. 2, pp. 552–559, 2022.). “K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon,”
- [9] Tajunisha and Saravanan, (vol. 1, no. 4, pp. 44–52, 2010). “Performance analysis of k-means with different initialization methods for high dimensional data,” *Int. J. Artif. Intell. Appl.*, doi: 10.5121/ijaia.2010.1404.
- [10] B. Harahap, (2019) “Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung),” *Reg. Dev. Ind. Heal. Sci. Technol. Art Life*, pp. 394–403 , [Online]. Available: <https://ptki.ac.id/jurnal/index.php/readystar/article/view/82>