

Klasterisasi Jumlah Tindak Pidana Kepolisian Daerah Pada Algoritma K-Means Klustering

¹ Adi Pangestu, ² Hairul Umam, ³ Florentina Wattilah, ⁴ Muhammad Gilang Ramadhan, ⁵ Sumanto, ⁶ Andi Diah Kuswanto
^{1,2,3,4,5,6} Universitas Bina Sarana Informatika

Alamat Surat

15220409@bsi.ac.id, hhoirull3322@gmail.com, florentinawattilah@gmail.com,
mhmmdgilang1511@gmail.com, sumanto.sto@bsi.ac.id, andi.ahk@bsi.ac.id

Article History:

Diajukan: 21 Juni 2025; Direvisi: 30 Juli 2025; Accepted: 29 November 2025

ABSTRAK

Penelitian ini bertujuan untuk menganalisis jumlah tindak pidana di Indonesia berdasarkan wilayah Kepolisian Daerah (Polda) dengan menerapkan algoritma K-Means Clustering. Data yang digunakan meliputi berbagai fitur numerik terkait tindak pidana dari sejumlah Polda di Indonesia. Proses analisis dilakukan dengan teknik normalisasi data serta pengelompokan menggunakan tiga klaster utama (C1, C2, C3), yang mencerminkan tingkat keparahan atau intensitas tindak pidana di masing-masing wilayah. Validitas hasil klaster diukur menggunakan nilai Silhouette, yang menunjukkan rata-rata sebesar 0,63, mengindikasikan bahwa pemisahan antar klaster cukup baik dan representatif. Hasil clustering memperlihatkan bahwa Polda-Polda yang tergolong dalam klaster C1 cenderung memiliki tingkat tindak pidana yang lebih rendah, ditunjukkan oleh nilai fitur yang relatif kecil. Sebaliknya, klaster C3 berisi wilayah dengan tingkat tindak pidana tertinggi, seperti Sumatera Utara dan beberapa wilayah padat penduduk lainnya. Klaster C2 berada di antara kedua kategori tersebut. Pengelompokan ini dapat membantu pihak berwenang dalam menyusun strategi penanggulangan kejahatan yang lebih tepat sasaran, dengan fokus pada klaster yang menunjukkan tingkat kejahatan tinggi. Penelitian ini menegaskan bahwa pendekatan data mining seperti K-Means Clustering efektif digunakan dalam menganalisis data kriminalitas dan memberikan insight geografis yang berguna untuk pengambilan keputusan. Ke depan, penelitian serupa dapat ditingkatkan dengan menambahkan variabel sosio-ekonomi atau demografis untuk memperkaya pemodelan dan interpretasi hasil.

Kata kunci: Analisis Kejahatan, Pengelompokan K-Means, Kepolisian Daerah, Penambahan Data, Tindak Pidana

ABSTRACT

This study aims to analyze the number of criminal offenses across regional police jurisdictions (Kepolisian Daerah) in Indonesia using the K-Means Clustering algorithm. The dataset includes normalized numerical features representing criminal activity metrics from various police regions. The clustering process utilized three main clusters (C1, C2, and C3), which represent varying levels of criminal activity intensity. The clustering validity was evaluated using the Silhouette coefficient, averaging at 0.63, indicating a reasonably good separation between clusters. The results show that police regions classified in cluster C1 tend to have lower levels of criminal offenses, as reflected by their smaller feature values. In contrast, cluster C3 includes areas with the highest levels of reported crime, such as North Sumatra and other densely populated regions. Cluster C2 falls in between, indicating moderate crime levels. This grouping can assist authorities in formulating more targeted

and region-specific crime prevention strategies, especially for areas within high-crime clusters. The findings demonstrate that data mining techniques such as K-Means Clustering can be effectively applied to analyze crime data and provide valuable geographical insights to support policy-making. Future research can enhance this model by incorporating additional socio-economic or demographic variables for deeper interpretation and improved predictive power.

Keywords: *Crime Analysis, K-Means Clustering, Regional Police, Data Mining, Criminal Offenses*

1. PENDAHULUAN

Tindak pidana merupakan permasalahan sosial yang kompleks dan terus berkembang, memengaruhi stabilitas dan keamanan masyarakat. Peningkatan jumlah kejahatan di berbagai wilayah Indonesia menuntut pendekatan analitis yang lebih canggih untuk memahami pola dan distribusinya. Dalam konteks ini, teknik data mining, khususnya algoritma K-Means Clustering, telah diterapkan secara luas guna mengelompokkan data kriminal berdasarkan karakteristik tertentu, memungkinkan identifikasi area rawan kejahatan secara lebih efektif. Metode K-Means termasuk dalam teknik unsupervised learning yang banyak digunakan dalam pengelompokan data kriminal, karena mampu mengidentifikasi pola tanpa memerlukan informasi awal yang rinci (Maharrani dkk.2024).

Penerapan algoritma K-Means yang dipadukan dengan teknologi Sistem Informasi Geografis (SIG) telah terbukti efektif dalam memetakan daerah rawan tindak pidana dan kemacetan. Sebagai contoh, suatu studi di Daerah Istimewa Yogyakarta menggunakan metode ini untuk memetakan pola kriminalitas ke dalam tiga klaster kerawanan (rendah, sedang, tinggi), menghasilkan visualisasi peta yang intuitif dan dapat langsung digunakan oleh aparat keamanan dalam perumusan strategi pengawasan wilayah (Putri dkk.2025).

Setelah klasterisasi menggunakan algoritma K-Means dan visualisasi melalui widget Scatter Plot di Orange, terlihat bahwa provinsi dikelompokkan berdasarkan skor Silhouette yang mencerminkan tingkat kohesi klaster dan jarak antar kelompok. Sebagai contoh, Polda Bengkulu masuk ke dalam Cluster C1 (tingkat tindak pidana rendah) dengan nilai siluet sekitar 0,63, sedangkan Polda Metro Jaya (Jakarta) tergolong dalam Cluster C3 (tingkat tindak pidana tinggi), dengan pewarnaan yang lebih terang menunjukkan kualitas klaster yang kuat. Temuan ini konsisten dengan studi lokal seperti penelitian oleh Anggoro et al. (2019) di Kota Semarang, yang menggunakan K-Means untuk memetakan daerah rawan pencurian kendaraan bermotor dan menghasilkan tiga klaster dengan pengelompokan spasial yang memadai (Wisma Anggoro dkk.2019).

Selain itu, studi oleh Aswan et al. (2021) di Kepulauan Mentawai juga menunjukkan bahwa penggunaan K-Means untuk klasifikasi area rawan kriminalitas berhasil mengidentifikasi wilayah dengan kerentanan tinggi, seperti Kecamatan Sipora Selatan, menegaskan validitas metode ini dalam pemetaan pola kejahatan (Aswan dkk.2021).

Selain itu, pendekatan pengelompokan ini juga telah diterapkan untuk mengklasifikasikan korban kejahatan di berbagai provinsi di Indonesia, membantu dalam memahami distribusi dan karakteristik korban secara lebih terstruktur. Penggunaan algoritma K-Means dan AHC untuk klasifikasi korban kejahatan berdasarkan lokasi geografis memungkinkan pemerintah menyusun kebijakan yang lebih efisien dalam distribusi sumber daya dan penanganan kasus (Lubis dkk.2023).

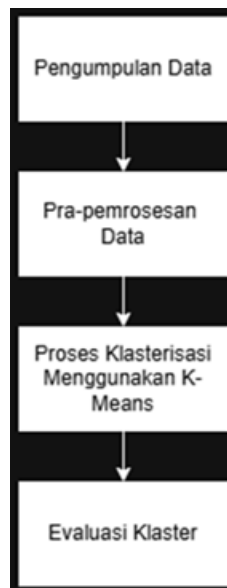
Latar belakang yang telah dipaparkan menjadi dasar bagi penelitian ini dalam melakukan analisis mengenai jumlah tindak pidana berdasarkan data Kepolisian Daerah (Polda) di Indonesia menggunakan algoritma K-Means Clustering. Dengan pendekatan tersebut, diharapkan mampu memberikan pemahaman yang lebih mendalam terkait pola distribusi kejahatan di berbagai wilayah, sehingga dapat mendukung perumusan kebijakan keamanan yang lebih efektif dan tepat sasaran.

2. METODE

Tahapan penelitian dilakukan melalui gambar sebagai berikut :

2.1 Metode Pengumpulan data

Sumber data dalam penelitian ini berasal dari situs resmi BPS, Terkhusus data jumlah tindak pidana yang tercatat pada masing masing kepolisian daerah di provinsi Indonesia. Data ini yang kemudian disusun kedalam format tabel untuk memudahkan proses analisis



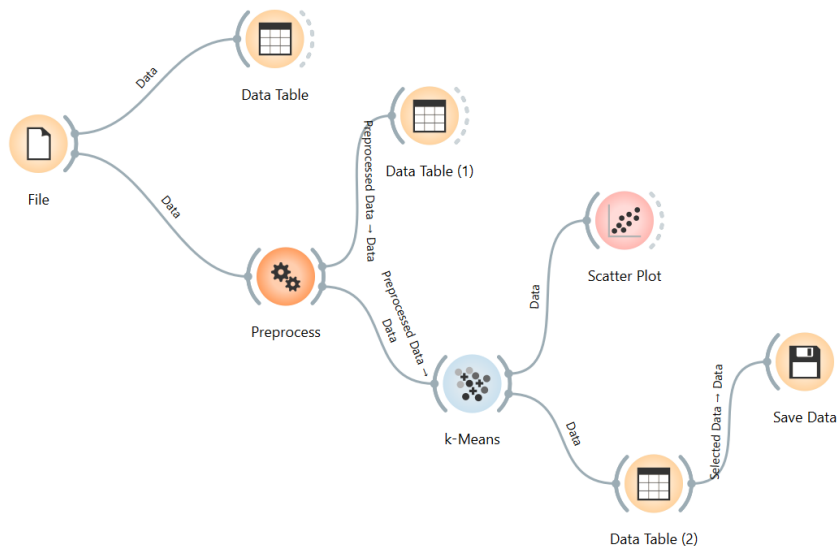
Gambar 1. Struktur Metode Penelitian

2.2 Pra-Pemrosesan Data

Sebelum data dilakukan analisis terlebih dahulu melalui proses pembersihan untuk menghilangkan nilai-nilai yang kosong atau tidak relevan. Kemudian, dilakukan normalisasi data tersebut menggunakan metode min-max normalization agar setiap variabel berada pada bagian skala yang sama. Data yang dinormalisasi dipakai untuk mencegah outlier dan mendapatkan kontribusi relatif yang lebih homogen antar variabel, sehingga hasil pengelompokan menjadi lebih tepat dan akurat karena pengelompokan objek didasarkan pada jarak antar titik data dan titik pusat (centroid) (Silvi, 2018).

2.3 Klasterisasi Menggunakan K-means

Data yang telah dinormalisasi kemudian dilakukan analisis menggunakan metode k-means clustering. Algoritma ini membagi data kedalam sejumlah klaster berdasarkan kemiripan nilai numerik antar wilayah provinsi, proses ini dilakukan dengan bantuan analisis perangkat lunak data yang memungkinkan visualiasi dan evaluasi hasil klastering.



Gambar 2. Workflow lengkap

2.4 Evaluasi kluster

Untuk mengevaluasi hasil dari klusterisasi digunakan silhouette coefficient, yang memberikan nilai seberapa baik setiap objek cocok dengan kluster sendiri dibandingkan dengan yang lain. Nilai ini menunjukkan koherensi internal dan pemisahan antar-kluster: semakin tinggi nilai rata-rata siluet semakin baik nilai kluster yang terbentuk, Dalam penelitian (Hidayati dkk.2021). Perbandingan antara beberapa metrik jarak jarak (misalnya Euclidean, Manhattan, dan Chebyshev) menunjukkan bahwa nilai silhouette score lebih dari 0,5 menunjukkan hasil klusterisasi yang baik dalam data dengan jumlah atribut yang bervariasi. Hasil ini mendukung penggunaan evaluasi siluet dalam penelitian ini, dimana nilai rata rata cluster-seperti aceh dengan skor sekitar 0,67-menggambarkan bahwa provinsi tersebut berada dalam kluster yang kohesif dan dipisahkan dengan baik dari kluster yang lainnya.

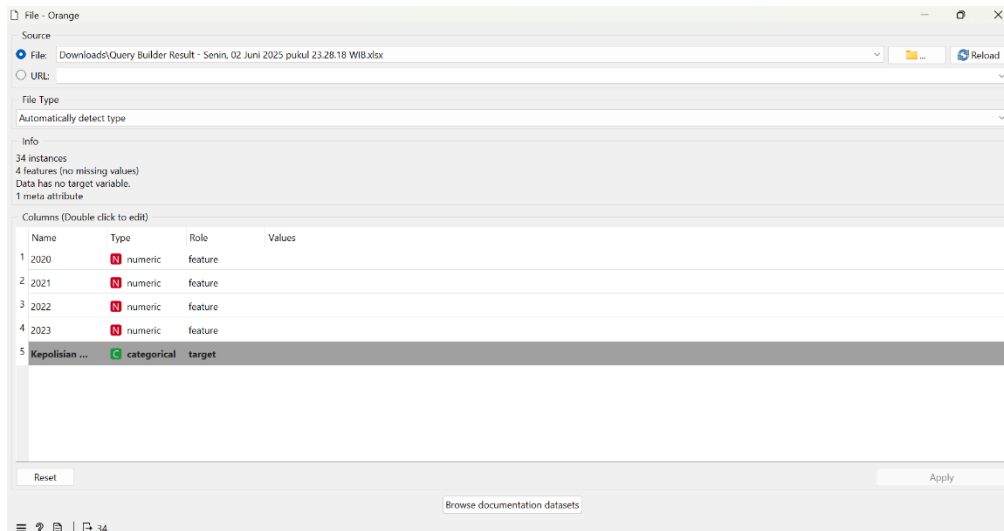
2.2 Tahapan Penelitian

Tahapan penelitian antara lain :

Tabel 1. Dataset

Kepolisian daerah	2020	2021	2022	2023
Aceh	7745	6651	10137	12420
Sumatera utara	32990	36534	43555	62278
Sumatera barat	7992	5666	7691	12722
Riau	8194	7512	12389	15777
Jambi	4709	3701	5359	7432

Sumatera selatan	12189	13037	11453	21335
Bengkulu	3333	3493	3613	5579
Lampung	7594	9764	11022	16608
Kep. Bangka belitung	1931	1566	2072	2211
Kep. Riau	2843	2481	3358	5074
Metro jaya	26585	29103	32534	87426
Jawa barat	11256	7502	29485	45694
Jawa tengah	10712	8909	30060	42304
Di yogyakarta	7721	4774	10591	12061
Jawa timur	17642	19257	51905	66741
Banten	4250	3434	5038	7392
Bali	2597	2404	6304	11916
Nusa tenggara barat	8591	6296	5296	7550
Nusa tenggara timur	4790	4909	5991	12692
Kalimantan barat	3858	4048	3975	6028
Kalimantan tengah	2629	2399	3189	4420
Kalimantan selatan	5206	4973	5016	6375
Kalimantan timur	3609	4564	4221	6762
Kalimantan utara	1015	971	1280	1701
Sulawesi utara	6274	6215	9618	14265
Sulawesi tengah	5454	5139	5453	8944
Sulawesi selatan	12815	14636	28679	41196
Sulawesi tenggara	2148	2431	3828	6276
Gorontalo	2518	2445	2488	3574
Sulawesi barat	1704	1500	2027	2679
Maluku	5350	3139	2383	4741
Maluku utara	850	1008	1220	2334
Papua barat	3162	2784	4083	6410
Papua	6962	6236	7584	14074



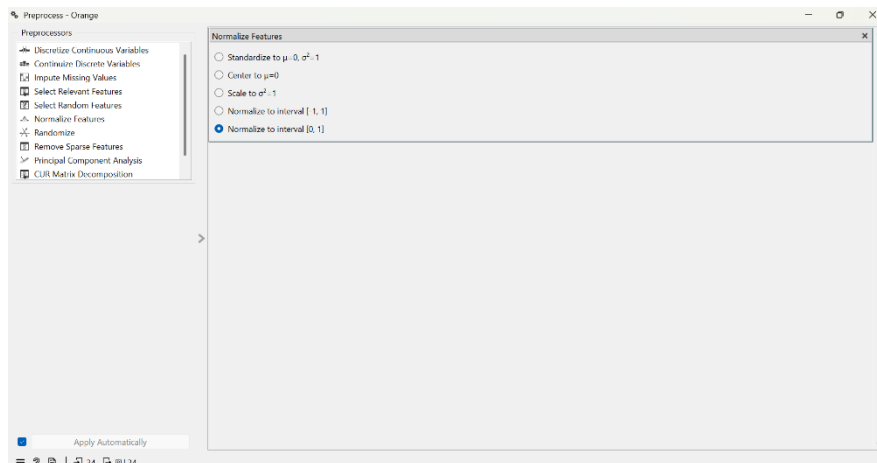
Gambar 3. File

Penggunaan widget pada file di Orange digunakan untuk memuat dataset yang berisi 34 instance dan 4 fitur numerik yaitu kolom tahun antara tahun 2020-2023. Sementara itu kolom kepolisian daerah telah terpilih sebagai target kategorikal walaupun masih ditandai sebagai meta atribut, sehingga langkah “Apply” dijalankan. Hal ini sejalan dengan prinsip data mining, dimana data input atau fitur dipisahkan dari data output yaitu target untuk mendukung model supervised learning. Studi sebelumnya oleh (Hartanto, n.d.). Bahwa penggunaan orange data mining memungkinkan pemodelan klasifikasi, termasuk juga pra-pemrosesan data validasi, seperti penanganan missing values dan pemilihan atribut yang tepat, untuk setelahnya diolah oleh berbagai algoritma seperti Decision Tree dan Naïve Bayes.

	Kepolisian daerah	2020	2021	2022	2023
1	ACEH	7745	6651	10137	12420
2	SUMATERA UT...	32990	36534	43555	62278
3	SUMATERA BA...	7992	5666	7691	12722
4	RIAU	8194	7512	12389	15777
5	JAMBI	4709	3701	5359	7482
6	SUMATERA SEL...	12189	13037	11453	21335
7	BENGKULU	3333	3493	3613	5579
8	LAMPUNG	7594	9764	11022	16608
9	KEP. BANGKA B...	1931	1566	2072	2211
10	KEP. RIAU	2843	2481	3358	5074
11	METRO JAVA	26585	29103	32534	87426
12	JAWA BARAT	11256	7502	29485	45694
13	JAWA TENGGAH	10712	8909	30060	42304
14	DI YOGYAKARTA	7721	4774	10591	12061
15	JAWA TIMUR	17642	19257	51905	66741
16	BANTEN	4250	3434	5038	7392
17	BAU	2597	2404	6304	11916
18	NUSA TENGAH...	8591	6296	5296	7550
19	NUSA TENGAH...	4790	4909	5991	12692
20	KALIMANTAN B...	3858	4048	3975	6028
21	KALIMANTAN T...	2629	2399	3189	4420
22	KALIMANTAN S...	5206	4973	5016	6375
23	KALIMANTAN T...	3609	4564	4271	6762
24	KALIMANTAN ...	1015	971	1280	1701
25	KALIMANTAN ...	6774	6715	8619	13765

Gambar 4. Tabel Mentah

Tampilan pada data table sangat berperan penting dalam tahap pertama proses data mining karena memungkinkan pengguna untuk melakukan validasi visual terhadap dataset sebelum lanjut ke tahap pemodelan. Dalam konteks ini, data table menunjukkan informasi penting seperti jumlah data instansi, fitur numerik pertahun (2020-2023), serta kolom kategori yang dijadikan target klasifikasi misalnya, “Kepolisian daerah”. Dari widget ini, pengguna dapat memastikan bahwa tidak ada nilai kosong (missing values), serta memverifikasi peran setiap atribut (feature atau target). Validasi ini adalah langkah untuk menghindari kesalahan dalam pemrosesan data. Kesesuaian metode ini juga ditunjukkan oleh (Pratama & Septian, 2023) yang menegaskan bahwa penggunaan Orange Data mining dalam pengklasifikasian kepribadian siswa yang menggunakan algoritma Acak Forest dan Naïve Bayes sangat bergantung kepada proses pra-pemrosesan dan validasi awal data lewat Data Table, sehingga menjamin keakuratan dan efektivitas model klasifikasi yang diterapkan.



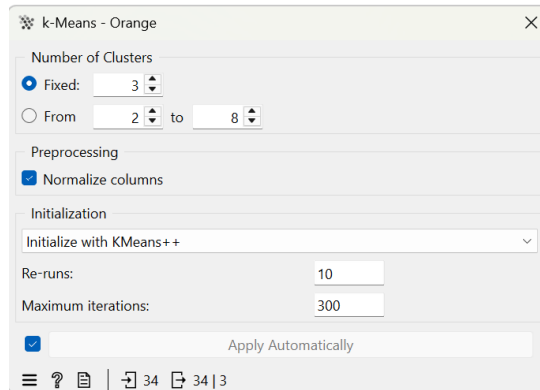
Gambar 5. Preprocess

Widget preprocess di Orange menyediakan berbagai opsi untuk normalisasi fitur numerik, salah satunya yaitu Normalize to interval [0.1] atau rescaling min-max. Normalisasi ini penting guna memastikan data berada pada skala yang beragam, sehingga variabel dengan rentang besar tidak mendominasi perhitungan yang sangat penting saat memakai algoritma klasifikasi berbasis jarak atau ensemble. [0,1] dipilih menunjukkan bahwa data akan dirubah ke rentang antara 0 hingga 1. Pendekatan ini mirip dengan yang dilakukan dalam penelitian (Oktaviani & Bahtiar, 2023), dimana normalisasi fitur sebelum input ke algoritma random forest terbukti meningkatkan akurasi klasifikasi pada data penjualan menggunakan aplikasi Orange.

	Kepolisian daerah	2020	2021	2022	2023
1	ACEH	0.2745302	0.1597186	0.1755298	0.1250394
2	SUMATERA UT.	1.00000	1.00000	0.832579	0.706643
3	SUMATERA BA.	0.2222133	0.1202192	0.1276709	0.1285623
4	RIAU	0.2395053	0.1630771	0.2203511	0.1641995
5	JAWA	0.1760885	0.0763959	0.0616619	0.066813
6	SUMATERA SEL.	0.3528003	0.338282	0.2019841	0.2380347
7	DIYOGAKARTA	0.0772308	0.0709164	0.042112	0.0432177
8	LAMPUNG	0.2098320	0.2472514	0.1933305	0.1738933
9	SEP. BANGKA B.	0.0336341	0.0767309	0.0168097	0.0059493
10	SEP. RIAU	0.0670100	0.0474598	0.0471871	0.0318467
11	METRO JAWA	0.0007156	0.7915889	0.6170159	1.00000
12	JAWA BARAT	0.3247710	0.1836459	0.1076401	0.1181795
13	JAWA TENGAH	0.3068511	0.2232055	0.1600046	0.1736425
14	DIYOGAKARTA	0.2137834	0.1065370	0.1646870	0.1208516
15	DIYOGAKARTA	0.5746447	0.5141861	1.00000	0.7587959
16	JAWA TENGAH	0.1057872	0.0692574	0.0753280	0.0633867
17	DIYOGAKARTA	0.0543559	0.0402947	0.1003358	0.1197601
18	DIYOGAKARTA	0.2496525	0.1497345	0.0804163	0.0632298
19	DIYOGAKARTA	0.1225887	0.1107331	0.0945304	0.1282123
20	KALIMANTAN B.	0.0619805	0.0605295	0.0545553	0.0504754
21	KALIMANTAN T.	0.0533516	0.0401541	0.0338078	0.0317177
22	KALIMANTAN S.	0.1355320	0.1725327	0.0748840	0.0546232
23	KALIMANTAN T.	0.0584532	0.1010320	0.0592088	0.0590376
24	KALIMANTAN	0.0051338	0.00000	0.0011838	0.00000

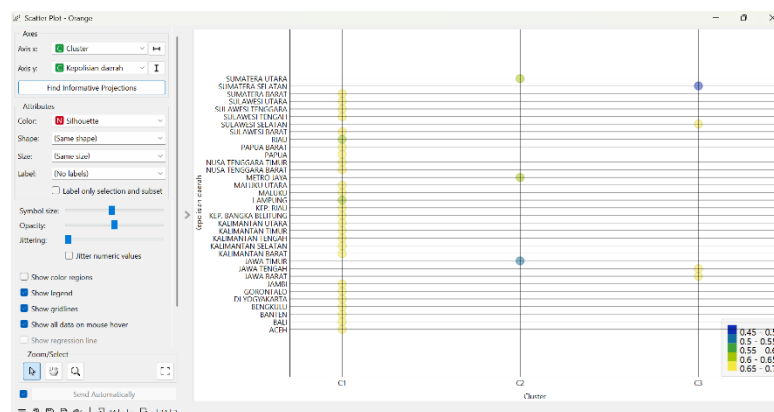
Gambar 6. Data Ternormalisasi

Setelah dilakukan proses normalisasi menggunakan cara metode min-max terhdap fiur numerik yaitu kolom tahun 2020-2023, widget data table menampilkan nilai-nilai yang sudah dikonversi menjadi skala 0-1. Proses ini bertujuan guna menyamakan skala antar fitur, sehingga tidak ada atribut yang mendominasi dari hasil perhitungan karena perbedaan rentan nilai. Penelitian oleh (Palinggik Allorerung dkk.2024), menunjukkan bahwa penggunaan metode normalisasi, termasuk Min-Max secara signifikan memengaruhi akurasi algoritma dalam klasifikasi dataset penyakit. Hal ini menegaskan bahwa transformasi nilai atribut merupakan proses penting dalam efektivitas algoritma klasifikasi.



Gambar 7. K-Means

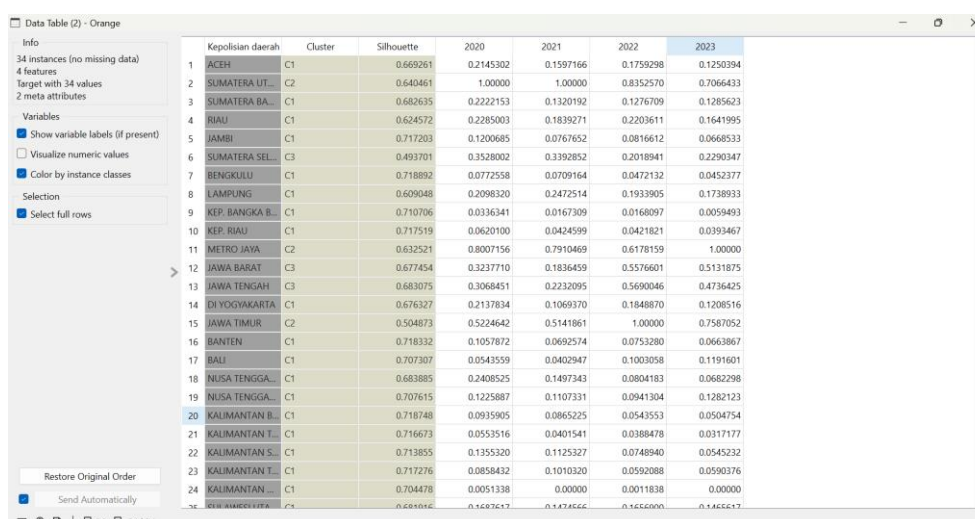
Tampilan K-Means tersebut mengelompokkan data ke dalam sejumlah klaster yang digunakan yaitu Fixed = 3 atau rentang 2-8 dengan opsi normlize columns aktif, penggunaan inisialisasi k-Means++, serta parameter Re-runs dan Maximum iterations untuk mencapai hasil yang optimal, normalisasi sangat penting karena K-means menggunakan distance metric sehingga fitur dengan rentang berbeda dapat menyebabkan klaster yang bias. Menurut penelitian (Yunita dkk.2024), penerapan normalisasi Z-Score sebelum K-Means mampu menghasilkan pengelompokan kasus UKT mahasiswa ini mengonfirmasi bahwa langkah pra-pemrosesan data seperti normalisasi fitur adalah bagian krusial dalam pipeline clustering yang dianjurkan dalam Aplikasi Orange data mining.



Gambar 8. Scatterplot

Visualisasi Hasil klasterisasi memakai widget scatter plot memperlihatkan distribusi masing-masing kepolisian daerah dalam tiga kelompok (C1,C2,C3) berdasarkan kemiripan jumlah tindak pidana. Sumbu x mempresentasikan nomor klaster, sedangkan sumbu y menampilkan nama polda. Setiap titik mewakili satu provinsi, dengan pewarnaan berdasarkan silhouette coefficient, yaitu metrik evaluasi internal yang mengukur seberapa baik suatu objek dengan klasternya di banding dengan klster yang lain. Warna yang lebih cerah menunjukan nilai siluet yang lebih tinggi,menandakan pemisahan klaster yang lebih baik.

Sebagai contoh polda bengkulu tergolong klaster C1 yang mewakili kelompok dengan tingkat tindak pidana rendah, dengan nilai 0,63. Sementara itu polda metro jaya masuk ke dalam cluster C3 yang menandakan wilayah dengan tindak pidana tinggi, juga dengan nilai siluet yang tinggi menandakan kejelasan dalam pemisahan antar kelompok. Penelitian sebelumnya menunjukkan bahwa nilai siluet 0,6 umumnya menunjukkan struktur klaster yang kuat dan terpisah dengan baik (Paembonan dkk. 2021). Pendekatan ini turut didukung terhadap temuan dalam penelitian yang dilakukan oleh (Aprianti & Permadi, 2018).dimana kombinasi algoritma K- Means dan visualiasi scatter plot digunakan secara efektif untuk mengelompokkan wilayah berdasarkan tingkat kecelakaan lalu lintas. Hasil studi mereka menunjukkan bahwa pemanfaatan metrik evaluasi seperti siluet mampu memberikan gambaran yang lebih objektif terhadap kualitas hasil pengelompokan.



Kepolisian daerah	Cluster	Silhouette	2020	2021	2022	2023
1 ACEH	C1	0.669261	0.2145302	0.1597166	0.1759298	0.1250394
2 SUMATERA UT.	C2	0.640461	1.00000	1.00000	0.8352570	0.7066433
3 SUMATERA BA.	C1	0.682635	0.2222153	0.1320192	0.1276709	0.1285623
4 RIAU	C1	0.624572	0.2285003	0.1839271	0.2203611	0.1641995
5 JAMBI	C1	0.717203	0.1200685	0.0767652	0.0816612	0.0668533
6 SUMATERA SEL.	C3	0.493701	0.3528002	0.3392852	0.2018941	0.2290347
7 BENGKULU	C1	0.718892	0.0772558	0.0709164	0.0472132	0.0452377
8 LAMPUNG	C1	0.609048	0.2098320	0.2472514	0.1933905	0.1738933
9 KEP. BANGKA B.	C1	0.710706	0.0336341	0.0167309	0.0168097	0.0059493
10 KEP. RIAU	C1	0.717519	0.0620100	0.0424599	0.0421821	0.0393467
11 METRO JAYA	C2	0.632521	0.8007156	0.7910469	0.6178159	1.00000
12 JAWA BARAT	C3	0.677454	0.3237710	0.1836459	0.5576601	0.5131875
13 JAWA TENGAH	C3	0.683075	0.3068451	0.2232095	0.5690046	0.4736425
14 DI YOGYAKARTA	C1	0.676327	0.2137834	0.1069370	0.1848870	0.1208516
15 JAWA TIMUR	C2	0.504873	0.5224642	0.5141861	1.00000	0.7587052
16 BANTEN	C1	0.718332	0.1057872	0.0692574	0.0753280	0.0663867
17 BALI	C1	0.707307	0.0543559	0.0402947	0.1003058	0.1191601
18 NUSA TENGA.	C1	0.683885	0.2408525	0.1497343	0.0804183	0.0682298
19 NUSA TENGA.	C1	0.707615	0.1225887	0.1107331	0.0941304	0.1282123
20 KALIMANTAN B.	C1	0.718748	0.0935905	0.0865225	0.0543553	0.0504754
21 KALIMANTAN T.	C1	0.716673	0.0553516	0.0401541	0.0368478	0.0317177
22 KALIMANTAN S.	C1	0.713855	0.1355320	0.1125327	0.0748940	0.0545232
23 KALIMANTAN T.	C1	0.717276	0.0858432	0.1010320	0.0592088	0.0590376
24 KALIMANTAN ...	C1	0.704478	0.0051338	0.00000	0.0011838	0.00000
25 ...	C1	0.681062	0.1687617	0.1474656	0.1656000	0.1465517

Gambar 9. Tabel Klasterisasi

Setelah proses klasterisasi memakai algoritma K-means, tampilan data table akan menampilkan kolom cluster (C1,C2,C3) dan silhouette, menunjukkan seberapa baik setiap daerah atau provinsi sesuai dengan klasternya tersendiri. Misalnya, provinsi Aceh berada pada di klaster C1 dengan skor silhouette 0,67, menandakan kohesi daerah dengan tindak pidana relatif rendah seperti Aceh, Jambi, dan Kalimantan barat. Sebaliknya, klaster C3, yang mencakup provinsi seperti sumatera selatan atau jawa barat, menunjukkan nilai yang lebih tinggi. Rata-rata silhouette 0,63 mengindikasikan pemisahan klaster yang cukup baik nilai ini sejalan dengan studi. yang melaporkan nilai siluet 0,73 saat menentukan jumlah klaster optimal pada analisis data kriminal (Sagala & Gunawan, 2022). Metode yang sama juga digunakan dalam studi DBSCAN pada kasus kriminalitas di Indonesia 2022, dengan nilai siluet mencapai 0,68 untuk memetakan daerah berdasarkan tingkat kejahatan (Miftahurrahmi dkk.2024).

3. HASIL DAN PEMBAHASAN

Proses klasterisasi yang menggunakan algoritma K-means telah teruji berhasil mengelompokkan 34 kepolisian daerah (POLDA) di Indonesia kedalam tiga klaster yaitu C1,C2,C3. Berdasarkan jumlah tindak pidana dari tahun 2020 sampai 2023. Data yang telah dinormalisasi menggunakan min-max normalization kemudian dianalisis melalui software Orange data mining.

Visualisasi scatter plot menunjukkan distribusi tiap polda berdasarkan kelompok klasternya, Rata-rata nilai silhouette coefficient yang diperoleh adalah 0,63, menunjukkan bahwa klaster yang terbentuk memiliki tingkat pemisahan yang jelas serta kohesi yang memadai. Hasil ini diperkuat dengan fakta bahwa polda seperti aceh memiliki nilai silhouette sebesar 0,67, menandakan kekompakan dalam klasternya. Adapun visualisasi pewarnaan berdasarkan silhouette dalam scatter plot menunjukkan bahwa pembagian klaster memiliki struktur yang jelas, warna yang terang, menandai klaster yang responsif.

4. SIMPULAN

Hasil penelitian ini membuktikan bahwa algoritma K-means clustering mampu digunakan secara efektif untuk menganalisis distribusi jumlah tindak pidana berdasarkan wilayah kepolisian di setiap daerah (provinsi) di Indonesia. Dengan membagi data kedalam tiga klaster utama, hasil analisis memberikan hasil informasi yang berguna mengenai tingkat kriminalitas di tiap wilayah di Indonesia.

Klaster C1 : adalah wilayah dengan tingkat pidana rendah

Klaster C2 : adalah wilayah dengan tingkat pidana sedang

Klaster C3 : adalah wilayah dengan tindak pidana tinggi

Dengan rata-rata nilai silhouette sebesar 0,63 menunjukkan bahwa klaster yang terbentuk memiliki kualitas yang cukup baik. Penemuan ini dapat dimanfaatkan oleh pihak berwenang sebagai dasar dalam merancang strategi pengendalian dan pencegahan kejahatan yang lebih spesifik, efektif, dan terarah.

5. DAFTAR PUSTAKA

- Aprianti, W., & Permadi, J. (2018). *K-MEANS CLUSTERING UNTUK DATA KECELAKAAN LALU LINTAS JALAN RAYA DI KECAMATAN PELAIHARI*. 5(5), 613–620. <https://doi.org/10.25126/jtiik2018551113>
- Aswan, Y., Defit, S., & Nurcahyo, G. W. (2021). Algoritma K-Means Clustering dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai). *Jurnal Sistim Informasi Dan Teknologi*, 245–250. <https://doi.org/10.37034/jsisfotek.v3i4.73>
- Hartanto, A. (n.d.). Implementasi Orange Data Mining Untuk Prediksi Penderita Diabetes. In *Prosiding Seminar Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan PROKASDADIK* (Vol. 1).

- Hidayati, R., Zubair, A., Hidayat Pratama, A., Indana, L., Studi Sistem Informasi, P., & Teknologi Informasi, F. (2021). *Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering* (Vol. 20, Issue 2).
- Lubis, R. M. F., Huang, J.-P., Wang, P.-C., Damanik, N., Sitepu, A. C., & Simanullang, C. D. (2023). K-Means and AHC Methods for Classifying Crime Victims by Indonesian Provinces: A Comparative Analysis. *Building of Informatics, Technology and Science (BITS)*, 5(1). <https://doi.org/10.47065/bits.v5i1.3630>
- Maharrani, R. H., Abda'u, P. D., & Faiz, M. N. (2024). Clustering method for criminal crime acts using K-means and principal component analysis. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(1), 224–232. <https://doi.org/10.11591/ijeecs.v34.i1.pp224-232>
- Miftahurrahmi, S., Zilrahmi, Amalita, N., & Mukhti, T. O. (2024). Metode Density Based Spatial Clustering of Applications with Noise (DBSCAN) dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Kriminalitas Tahun 2022. *UNP Journal of Statistics and Data Science*, 2(3), 330–337. <https://doi.org/10.24036/ujsds/vol2-iss3/203>
- Oktaviani, S., & Bahtiar, A. (2023). Implementasi Algoritma K-Means Dalam Pengelompokan Data Penjualan CV. Widuri Menggunakan Orange. In *Jurnal Wahana Informatika (JWI)* (Vol. 2, Issue 1).
- Paembonan, S., Abduh, H., & Kunci, K. (2021). *Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coefficient* (Vol. 6, Issue 2). <https://ojs.unanda.ac.id/index.php/jiit/index>
- Palinggik Allorerung, P., Erna, A., Bagussahrir, M., & Alam, S. (2024). Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit. In *Jurnal Informatika Sunan Kalijaga* (Vol. 9, Issue 3).
- Pratama, & Septian. (2023). *referensi data table 1. Implementasi Orange Untuk Klasifikasi Kepribadian Siswa-Siswi SMAN 79 Jakarta Dengan Model Random Forest dan Naive Bayes*.
- Putri, A. E., Mz, Y., & Bororing, J. E. (2025). IMPLEMENTASI K-MEANS CLUSTERING DAN MODEL CRISP-DM UNTUK PENGELOMPOKAN DAERAH RAWAN TINDAK PIDANA NARKOBA DI DIY. *Jurnal Informatika Teknologi Dan Sains*, 7(2), 615.
- Sagala, N. T. M., & Gunawan, A. A. S. (2022). Discovering the Optimal Number of Crime Cluster Using Elbow, Silhouette, Gap Statistics, and NbClust Methods. *ComTech: Computer, Mathematics and Engineering Applications*, 13(1), 1–10. <https://doi.org/10.21512/comtech.v13i1.7270>
- Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia. *Jurnal Matematika "MANTIK,"* 4(1), 22–31. <https://doi.org/10.15642/mantik.2018.4.1.22-31>
- Wisma Anggoro, J., Awaluddin, M., & Laila Nugraha, A. (2019). ZONASI DAERAH RAWAN PENCURIAN KENDARAAN BERMOTOR (CURANMOR) DI KOTA SEMARANG DENGAN MENGGUNAKAN METODE CLUSTER ANALYSIS. In *Jurnal Geodesi Undip Oktober* (Vol. 8).
- Yunita, Y., Fahmi, M., & Salmon, S. (2024). Penerapan Algoritma K-Means Data Mining Pada Clustering Kelayakan Penerima UKT Dengan Normalisasi Data Model Z-Score. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1977–1986. <https://doi.org/10.47065/bits.v6i3.6475>