

PERBANDINGAN *K-MEANS* DAN *K-MEDOIDS* DALAM PENGELOMPOKAN TINGKAT KEJAHATAN PADA PROVINSI JAWA TENGAH

Bayu Satria Pratama^{1*}, Gatot Purwanto²

¹Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

²Sistem Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹bayupratama@budiluhur.ac.id, ²gatot.purwanto@budiluhur.ac.id

(* : corresponding author)

Abstrak- Tindak kejahatan merupakan pelanggaran hukum dan norma sosial yang menimbulkan keresahan masyarakat serta mencerminkan dinamika sosial yang terus berkembang. Sepanjang tahun 2023, Provinsi Jawa Tengah mencatat 7.606 kasus kejahatan berdasarkan data dari Badan Pusat Statistik (BPS) Jawa Tengah pada website <https://jateng.bps.go.id/id>. Jenis kejahatan yang tercakup dalam data merupakan kejahatan konvensional, seperti pencurian, penganiayaan, dan penipuan. Penelitian ini bertujuan untuk mengelompokkan data kriminalitas menggunakan algoritma clustering *K-Means* dan *K-Medoids* guna mengidentifikasi pola kejahatan berdasarkan karakteristik wilayah. Data yang digunakan meliputi jumlah kejahatan, jumlah penduduk, dan jumlah penduduk tidak bekerja per kabupaten/kota, serta dua atribut turunan, yaitu Rasio_Kejahatan_Penduduk dan Rasio_Kejahatan_Tidak Bekerja. Seluruh data numerik dinormalisasi menggunakan metode *Min-Max* agar memiliki skala yang sebanding. Pemilihan algoritma *K-Means* dan *K-Medoids* dilakukan karena keduanya merupakan metode *partitional clustering* yang banyak digunakan, namun memiliki pendekatan yang berbeda dalam menentukan pusat kluster, sehingga memberikan perbandingan hasil yang relevan. Evaluasi hasil kluster dilakukan dengan *Davies-Bouldin Index (DBI)* karena metrik ini mampu menilai validitas kluster berdasarkan tingkat kedekatan dan keterpisahan antar kluster. Klusterisasi dilakukan dengan jumlah kluster 3, 4, dan 5. Hasil evaluasi menunjukkan bahwa pembentukan 3 kluster adalah yang paling optimal, dengan nilai *DBI* terendah pada *K-Means* sebesar 0,082, sedikit lebih baik dibandingkan *K-Medoids* sebesar 0,084. Nilai *DBI* yang lebih rendah menunjukkan *K-Means* menghasilkan kluster yang lebih terpisah secara baik. Oleh karena itu, *K-Means* dipilih sebagai algoritma terbaik dalam penelitian ini. Hasil pengelompokan diharapkan menjadi dasar dalam persamaan kebijakan.

Kata Kunci: Algoritma *K-Means*, *Davies-Bouldin Index*, Kejahatan Konvensional, Klusterisasi, Provinsi Jawa Tengah

Abstract- *Criminal acts are violations of law and social norms that cause public unrest and reflect the continuously evolving dynamics of society. Throughout 2023, Central Java Province recorded 7,606 crime cases, based on data from the Central Java Statistics Agency (BPS) available on the website <https://jateng.bps.go.id/id>. Types of crimes included are conventional crimes such as theft, assault, and fraud. This study aims to cluster crime data using K-Means and K-Medoids clustering algorithms to identify crime patterns based on regional characteristics. Data used includes the number of crimes, total population, and number of unemployed people per district/city and Crime-to-Population Ratio and Crime-to-Unemployed Ratio. All numerical data were normalized using the Min-Max method to ensure comparable scales. The choice of K-Means and K-Medoids algorithms was made because both are widely used partitional clustering methods, yet they adopt different approaches in determining cluster centers, thus providing a relevant comparison of results. Cluster evaluation was conducted using the Davies-Bouldin Index (DBI), as this metric effectively assesses cluster validity based on the compactness and separation of clusters. Clustering was carried out with 3, 4, and 5 clusters. The evaluation results 3 clusters was the most optimal, with the lowest DBI value of 0.082 for K-Means, slightly better than 0.084 for K-Medoids. A lower DBI value indicates that K-Means produced better-separated clusters. Therefore, K-Means was selected as the best-performing algorithm. The results are expected to serve as a foundation for formulating more accurate crime prevention policies that align with the local characteristics of each region.*

Keywords: *Central Java Province, Clustering, Conventional Crime, Davies-Bouldin Index, K-Means Algorithm*

1. PENDAHULUAN

Tindak Kejahatan merupakan tindakan yang bertentangan dengan hukum dan norma sosial dikenal sebagai bentuk pelanggaran yang dapat memicu ketidaknyamanan serta keresahan dalam kehidupan masyarakat. Fenomena ini dipahami sebagai bagian dari dinamika sosial yang terus berkembang seiring dengan perubahan dalam struktur dan pola kehidupan masyarakat. Berbagai bentuk pelanggaran hukum yang terjadi dapat dikelompokkan ke dalam kategori tindak pidana umum dan khusus[1]. Sepanjang tahun 2023 tercatat sebanyak 7.606 kasus kejahatan yang terjadi pada Provinsi Jawa Tengah[2].

Angka kejahatan yang tinggi menunjukkan bahwa masalah ini perlu perhatian serius dari berbagai pihak. Faktor-faktor seperti kesenjangan sosial, rendahnya pendidikan, dan lemahnya penegakan hukum berkontribusi pada tingginya angka kriminalitas. Oleh karena itu, dibutuhkan pendekatan menyeluruh melalui pencegahan, pemberdayaan masyarakat, dan peningkatan kualitas layanan kepolisian. Salah satu langkah penting adalah melakukan klusterisasi

kasus kejahatan. Dengan klustering, penanggulangan kejahatan dapat lebih terfokus dan sesuai dengan kondisi lokal, sehingga kebijakan yang diambil akan lebih efektif dan tepat sasaran[1].

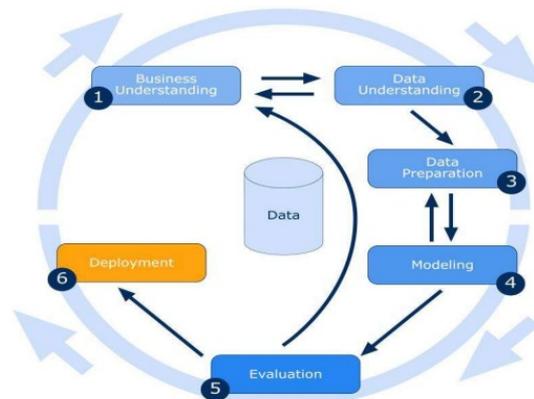
Salah satu teknik yang umum digunakan dalam membentuk kluster adalah *K-Means* dan *K-Medoids*[3]. *K-Means* adalah algoritma clustering yang membagi data ke dalam beberapa grup berdasarkan kemiripan antar objek. Setiap grup memiliki objek yang mirip satu sama lain, sementara objek di grup berbeda memiliki karakteristik yang berbeda[4]. *K-Medoids* menggunakan objek data itu sendiri sebagai pusat kluster (*medoid*). Algoritma ini lebih tahan terhadap noise dan outlier dibandingkan *K-Means*, karena *medoid* tidak terpengaruh oleh nilai ekstrem yang bisa mengganggu distribusi data[5].

Kedua teknik ini merupakan bagian dari metode *data mining* yang dirancang untuk mengelompokkan data berdasarkan pola atau karakteristik tertentu. Teknik ini dapat membantu menganalisis data dengan lebih efektif, sehingga strategi penanggulangan kejahatan dapat disesuaikan dengan kebutuhan spesifik di setiap wilayah[6], [7]. Beberapa penelitian terdahulu yang menggunakan Teknik ini dalam membentuk kluster pada tingkat kejahatan diantaranya Implementasi algoritma *K-Means* bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan sejumlah faktor yang berpengaruh terhadap tingkat tindak pidana[8]. Berikutnya ada analisis perbandingan algoritma *K-Means* dan *K-Medoids* dalam pengelompokan tipe kriminalitas[9].

Penerapan algoritma *K-Means* dan *K-Medoids* dapat dievaluasi untuk menentukan model *clustering* terbaik. Metode evaluasi yang digunakan adalah *DBI*, yaitu metrik yang mengukur seberapa baik hasil pengelompokan dengan menilai tingkat kesamaan dalam kluster dan perbedaan antar kluster[7], [10]. Tujuan dari penelitian ini adalah untuk menerapkan dan membandingkan algoritma *K-Means* dan *K-Medoids* dalam mengelompokkan data kriminalitas di Provinsi Jawa Tengah. Pengelompokan ini dilakukan untuk mengidentifikasi pola kejahatan berdasarkan karakteristik tertentu, sehingga dapat mendukung pepersamaan kebijakan penanggulangan yang lebih efektif. Evaluasi kualitas hasil clustering dilakukan menggunakan metode *Davies-Bouldin Index (DBI)* guna menentukan model terbaik. Hasil dari penelitian ini diharapkan memberikan rekomendasi model pengelompokan yang akurat dan sesuai dengan kondisi lokal, serta membantu pihak berwenang dalam upaya pencegahan dan pengendalian tindak kriminal.

2. METODE PENELITIAN

2.1. CRIPS-DM



Gambar 1. Metode CRIPS-DM

Untuk menyelesaikan masalah ini, digunakan metode *CRISP-DM*, dalam *data mining* *CRIPS-DM* merupakan kerangka standar dalam data mining yang dikembangkan oleh *Daimler Chrysler*, *SPSS*, dan *NCR*. Metode ini diilustrasikan dalam Gambar 1 dan terdiri dari 6 tahapan [7], [11], [12], yaitu:

- Business Understanding*
Tahap ini bertujuan untuk memahami tujuan bisnis secara menyeluruh dan mempersamakan masalah data mining. Keputusan di tahap ini akan menentukan arah seluruh proses selanjutnya.
- Data Understanding*
Tahap ini melibatkan pengumpulan data awal, eksplorasi data, dan identifikasi masalah kualitas data.
- Data Preparation*
Pada tahap ini, data dibersihkan, diintegrasikan, dipilih, dan diformat ke dalam bentuk yang sesuai untuk proses modeling. Ini adalah salah satu tahap paling memakan waktu dalam proses data mining.
- Modeling*
Di tahap ini, algoritma data mining diterapkan pada data yang telah disiapkan. Beberapa teknik modeling dapat dicoba, dan parameter model diatur untuk mendapatkan hasil terbaik.

e. *Evaluation*

Setelah model dibuat, langkah ini bertujuan untuk mengevaluasi apakah model yang dihasilkan benar-benar memenuhi tujuan bisnis. Evaluasi dilakukan menggunakan metrik yang relevan dan analisis interpretasi hasil.

f. *Deployment*

Tahap akhir ini mencakup penerapan model ke dalam lingkungan nyata atau sistem operasional maupun sistem otomatis yang dapat mendukung pengambilan keputusan.

2.2. K-Means

Algoritma *K-Means* merupakan metode yang biasa digunakan untuk membentuk model clustering dengan cara melihat kemiripan karakteristik data yang kemudian dikelompokkan. Sebagai metode pengelompokan data non-hierarki, *K-Means* membagi *dataset* ke dalam klaster-klaster yang berbeda, di mana data berkarakteristik sama dikelompokkan dalam klaster yang sama, sedangkan data dengan karakteristik yang berbeda akan ditempatkan pada klaster lain [13]–[15]. *K-Means* bekerja dengan cara menentukan jumlah dari klaster (k), menentukan *centroid* pertama, jarak *Euclidean* setiap data ke *centroid* dihitung, melakukan pengelompokan data berdasarkan jarak terdekat, lalu *centroid* sebagai rata-rata anggota klaster akan diperbaharui, formula *Euclidean Distace* ada pada persamaan (1) dan menghitung *centroid* baru sebagai rata-rata anggota klaster ada pada persamaan (2) dan persamaan untuk menentukan klaster pada persamaan (3) [16], [17].

$$d(x+y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$

$$C_k = \frac{1}{n_k} \sum d_i \tag{2}$$

$$c_j = \left(\frac{1}{n_j}\right) \sum_{x \in Cluster_j} x \tag{3}$$

2.3. K-Medoids

Algoritma *K-Medoids* menggunakan langkah inisialisasi *medoid* secara acak, mengelompokan data berdasarkan jarak Manhattan terdekat, lalu memperbarui *medoid* dengan memilih kandidat yang meminimalkan total jarak dalam klaster. Proses ini diulang hingga *medoid* stabil. Formula jarak Manhattan ada pada persamaan (4)[18].

$$d(a_x, b_y) = \sum_{z=1}^n |a_z - b_z| \tag{4}$$

2.4. Min-Max

Metode *Min-Max* adalah teknik normalisasi yang mentransformasi data asli secara linier sehingga nilainya berada dalam rentang 0 hingga 1. Proses ini menyesuaikan setiap nilai berdasarkan batas bawah dan atas dari fitur terkait. Nilai yang baru diperoleh dengan cara mengurangi nilai awal dengan batas bawah, kemudian dibagi dengan selisih antara batas atas dan batas bawah fitur tersebut. Formula *Min-Max* ada pada persamaan (5)[19].

$$x_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \tag{5}$$

2.5. Davies-Bouldin Index (DBI)

Evaluasi kualitas klaster akan diuji dengan *Davis-Bouldin Index (DBI)* yang merupakan metode untuk menentukan kualitas *clustering*. Nilai *DBI* yang kecil, maka kualitas kluster yang dihasilkan semakin baik. Jumlah kluster optimal ditentukan dari nilai *DBI* terkecil dengan persamaan pada persamaan (6)[14], [17].

$$DB = \left(\frac{1}{N}\right) \sum_{i=1}^N \max_{j \neq i} (R_{ij}) \tag{6}$$

2.6. Perhitungan Rasio

Perhitungan Perhitungan rasio dilakukan dengan membagi jumlah kejadian (frekuensi) dengan jumlah populasi (penduduk) pada wilayah atau kelompok tertentu. Secara matematis, rasio ini dapat dinyatakan dengan persamaan (7)[20].

$$P = \frac{F}{N} \quad (7)$$

P= Rasio

N= Penduduk/Tidak Bekerja

F= Jumlah Kejahatan

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding (Pemahaman Bisnis)

Tujuan utama dari penelitian adalah mengelompokkan data tindak kriminal di Provinsi Jawa Tengah agar pola kejahatan dapat teridentifikasi berdasarkan karakteristik tertentu. Hasil dari pengelompokan ini digunakan untuk mendukung perumusan kebijakan penanggulangan kejahatan yang lebih tepat sasaran dan efektif.

3.2. Data Understanding (Pemahaman Data)

Mengumpulkan serta mempelajari data kriminalitas yang terjadi di Provinsi Jawa Tengah. Data bisa lokasi kejadian, dan jumlah kasus, serta faktor-faktor sosial ekonomi lain yang memengaruhi tingkat kriminalitas. Pada tahap ini dilakukan eksplorasi awal data untuk melihat distribusi, kecenderungan, dan kemungkinan anomali. Data diperoleh <https://jateng.bps.go.id/id> yang merupakan situs resmi dari Badan Pusat Statistik Provinsi Jawa Tengah dan menyediakan informasi statistik terkait kriminalitas dan indikator sosial ekonomi lainnya.

3.3. Data Preparation (Persiapan Data)

Pada tahapan ini melibatkan pembersihan data (*data cleaning*), transformasi data, dan seleksi fitur yang relevan untuk proses *clustering*. Contoh kegiatan pada tahap ini meliputi normalisasi data, penghapusan data duplikat atau outlier, serta pemilihan atribut yang akan digunakan dalam analisis klaster. Bentuk data diilustrasikan pada Tabel 1.

Tabel 1. Data Kejahatan

Kabupaten/Kota	Jumlah_Kejahatan	Jumlah_Penduduk	Tidak_Bekerja
Cilacap	298	2.007.829	91.510
Banyumas	326	1.828.573	58.727
...
Kota Pekalongan	125	1.007.384	14.607
Tegal	159	1.654.836	73.171

Data menunjukkan kabupaten/kota yang ada di Provinsi Jawa Tengah memiliki tingkat kejahatan dan jumlah penduduk tidak bekerja yang bervariasi. Beberapa daerah dengan jumlah penduduk tinggi cenderung mencatat angka kejahatan dan pengangguran yang lebih besar dibandingkan daerah lain. Perbedaan ini mencerminkan adanya ketimpangan sosial yang dapat memengaruhi tingkat kriminalitas, sehingga penting untuk dianalisis lebih lanjut dalam upaya perumusan kebijakan yang tepat.

Pada tahap *data preparation*, akan ditambahkan dua atribut baru ke dalam data, yaitu rasio kejahatan terhadap jumlah penduduk yang dinamakan *Rasio_Kejahatan_Penduduk*, serta rasio kejahatan terhadap jumlah penduduk yang tidak bekerja dengan nama *Rasio_Kejahatan_Tidak_Bekerja*. Penambahan atribut-atribut ini dilakukan menggunakan *RapidMiner*. Penambahan atribut *Rasio_Kejahatan_Penduduk* dan *Rasio_Kejahatan_Tidak_Bekerja* dilakukan dengan menggunakan operator *Generate Attributes* pada *RapidMiner*. Operator ini digunakan untuk menambahkan atribut baru ke dalam data, serta menyisipkan formula perhitungan rasio melalui *function expressions*.

Tabel 2. Dataset tindak kejahatan setelah penambahan atribut

Kabupaten/ Kota	Jumlah_Kejahatan	Jumlah_Penduduk	Tidak_Bekerja	Rasio_Kejahatan_Penduduk	Rasio_Kejahatan_Tidak_Bekerja
Cilacap	298	2.007.829	91.510	0.000	0.003
Banyumas	326	1.828.573	58.727	0.000	0.006
...
Kota Pekalongan	125	1.007.384	14.607	0.000	0.007
Tegal	159	1.654.836	73.171	0.000	0.002

Setelah melalui proses Generate Attributes, data baru yang terbentuk terdapat pada Tabel 2. Data yang ada pada Tabel 2 mencerminkan hasil penambahan atribut berdasarkan perhitungan rasio, pada Rapidminer hanya bisa menampilkan maksimal 3 angka dibelakang koma. Tahap selanjutnya adalah melakukan normalisasi menggunakan metode *Min-Max*. Normalisasi *Min-Max* pada *RapidMiner* berfungsi untuk mengubah skala nilai atribut numerik ke dalam rentang tertentu, umumnya antara 0 hingga 1. Proses ini merupakan bagian penting dari tahap data preprocessing untuk meyakinkan bahwa seluruh atribut mempunyai skala yang seragam. Normalisasi juga bertujuan untuk mencegah dominasi atribut dengan nilai besar terhadap atribut lainnya dalam proses analisis. Normalisasi *Min-Max* pada *Rapidminer*. Hasil dari proses normalisasi *Min-Max* terlihat di Tabel 3.

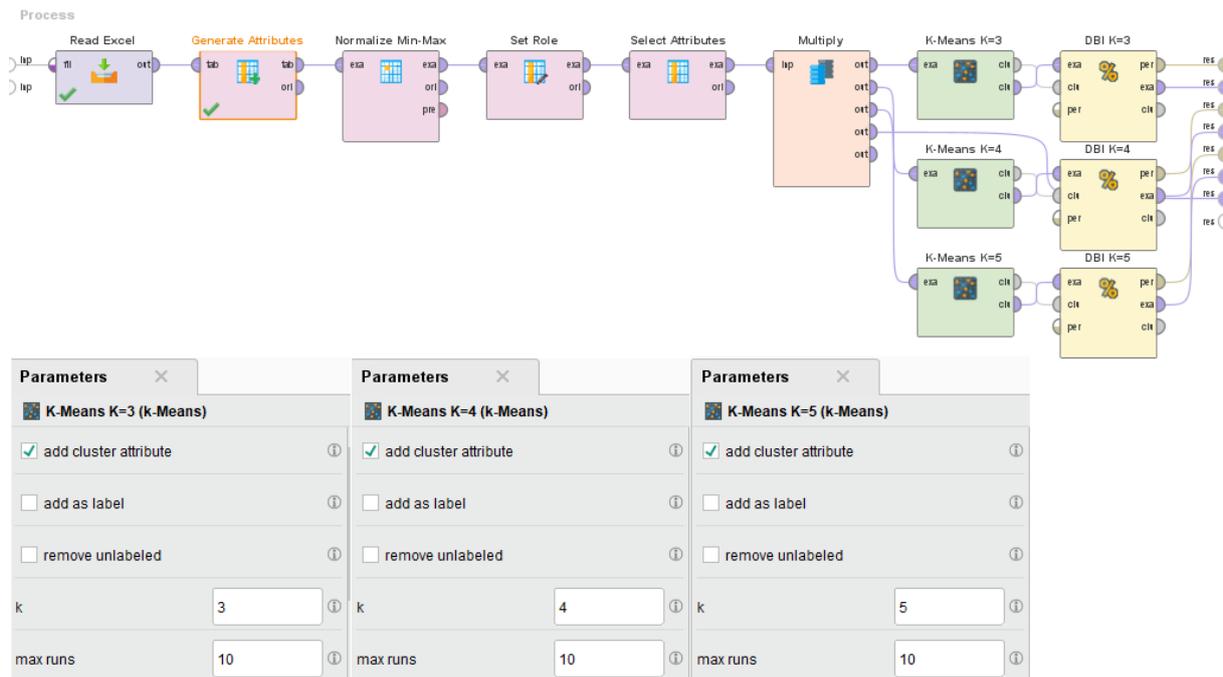
Tabel 3. Data Tindak Kejahatan setelah dilakukan normalisasi *Min-Max*

Kabupaten/ Kota	Jumlah_Kejahatan	Jumlah_Penduduk	Tidak_Bekerja	Rasio_Kejahatan_Penduduk	Rasio_Kejahatan_Tidak_Bekerja
Cilacap	0.180	0.981	0.979	0.070	0.026
Banyumas	0.203	0.884	0.607	0.096	0.075
...
Kota Pekalongan	0.037	0.438	0.141	0.048	0.107
Tegal	0.065	0.789	0.771	0.023	0.003

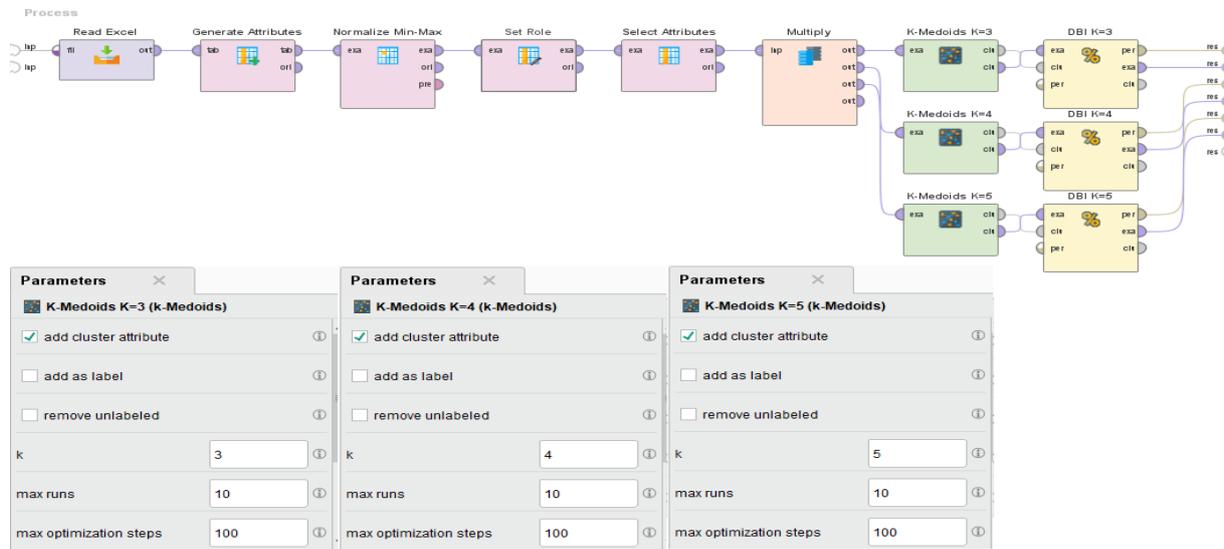
3.4. Modeling (Pemodelan)

Pada tahap *modeling*, algoritma *K-Means* dan *K-Medoids* diterapkan untuk melakukan proses *clustering* terhadap data kriminalitas di Provinsi Jawa Tengah. Algoritma ini mengelompokkan wilayah berdasarkan kemiripan karakteristik tindak kejahatan, sehingga pola-pola kejahatan yang tersebar di berbagai daerah dapat diidentifikasi dengan lebih mudah. *K-Means* bekerja dengan membentuk kluster dari data yang sudah dibagi, kluster akan terbentuk berdasarkan pusat gravitasi (*centroid*), sementara *K-Medoids* menggunakan objek aktual dari data sebagai pusat kluster (*medoid*), yang membuatnya lebih tahan terhadap *outlier* dan *noise*. Pemilihan algoritma ini bertujuan untuk membandingkan performa masing-masing metode dalam menghasilkan kluster yang representatif.

Jumlah kluster yang optimal tidak ditentukan secara langsung, melainkan melalui proses evaluasi yang mempertimbangkan tingkat kedekatan dalam kluster serta pemisahan antar kluster, menggunakan metrik tertentu seperti *Davies-Bouldin Index (DBI)*. Untuk merealisasikan implementasi algoritma tersebut, dilakukan perancangan atau penyusunan alur proses dengan menggunakan operator yang tersedia di platform *RapidMiner*. Platform ini menyediakan antarmuka visual berbasis blok yang memungkinkan pengguna merancang proses analitik secara sistematis dan efisien. Rangkaian operator untuk implementasi algoritma *K-Means* diilustrasikan Gambar 2, sedangkan rangkaian proses untuk algoritma *K-Medoids* ditampilkan pada Gambar 3.



Gambar 2. Model *K-Means*



Gambar 3. Model K-Medoids

3.5. Evaluation (Evaluasi)

Jumlah kluster yang terbentuk dalam proses *clustering* ditentukan sebanyak tiga variasi, yaitu 3, 4, dan 5 kluster. Pembagian ini diterapkan secara konsisten pada kedua algoritma, baik *K-Means* maupun *K-Medoids*, guna memperoleh perbandingan yang adil dalam hasil pengelompokan. Setelah tahap *modeling* selesai dan seluruh proses *clustering* berhasil dilakukan untuk masing-masing jumlah kluster, langkah selanjutnya adalah melakukan evaluasi terhadap kualitas kluster yang terbentuk. Evaluasi dilakukan dengan menggunakan metode *DBI*, yang berfungsi menilai seberapa baik kluster terbentuk berdasarkan tingkat kemiripan dalam kluster (*intra-cluster similarity*) dan perbedaan antar kluster (*inter-cluster separation*). Semakin rendah atau kecil nilai *DBI* memperlihatkan bahwa kluster yang terbentuk lebih kompak dan telah dipisah dengan baik, sehingga dapat dianggap sebagai hasil pengelompokan yang optimal. Evaluasi ini penting untuk menentukan jumlah kluster terbaik dan algoritma yang paling sesuai dengan karakteristik data kriminalitas yang dianalisis.

Setelah melalui proses evaluasi menggunakan metode *Davies-Bouldin Index (DBI)*, diperoleh nilai *DBI* untuk masing-masing jumlah kluster pada *K-Means*. Nilai tersebut menggambarkan kualitas pengelompokan berdasarkan tingkat kemiripan dalam kluster dan perbedaan antar kluster. Hasil evaluasi *DBI* untuk algoritma *K-Means* ditampilkan pada Tabel 4. Selanjutnya, evaluasi yang sama juga dilakukan terhadap hasil *clustering* yang diperoleh dari algoritma *K-Medoids*, dengan tujuan untuk membandingkan performa kedua metode dalam menghasilkan kluster yang optimal. Nilai *DBI* untuk masing-masing jumlah kluster pada *K-Medoids* dapat dilihat pada Tabel 4. Perbandingan hasil evaluasi ini menjadi dasar dalam menentukan algoritma dan jumlah kluster terbaik yang paling sesuai dengan pola distribusi data kriminalitas yang dianalisis.

Tabel 4. DBI K-Means dan K-Medoids

Jumlah Kluster	DBI K-Means	DBI K-Medoids
K=3	0.082	0.084
K=4	0.094	0.099
K=5	0.127	0.140

Tabel 4 menunjukkan nilai *DBI* untuk setiap jumlah kluster yang diterapkan pada kedua algoritma. Dari hasil evaluasi, terlihat bahwa pembagian menjadi 3 kluster memberikan nilai *DBI* terendah untuk kedua algoritma, yaitu 0,082 pada *K-Means* dan 0,084 pada *K-Medoids*. Nilai ini menunjukkan bahwa kluster yang terbentuk pada konfigurasi tiga kluster memiliki tingkat kedekatan antar data dalam satu kluster yang baik serta pemisahan antar kluster yang jelas. Sementara itu, ketika jumlah kluster ditambah menjadi 4 dan 5, nilai *DBI* cenderung meningkat, yang mengindikasikan bahwa kualitas kluster justru menurun. Misalnya, pada K=5, nilai *DBI* yang dimiliki *K-Means* mencapai 0,127 dan hasil *K-Medoids* sebesar 0,140. Berdasarkan hal tersebut, dapat disimpulkan bahwa jumlah kluster optimal untuk data kriminalitas yang dianalisis adalah 3 kluster, karena menghasilkan pengelompokan yang paling efisien dan terstruktur menurut indikator *DBI*.

Data hasil pengelompokan menggunakan *K-Means* dengan 3 kluster dan hasil pengelompokan menggunakan *K-Medoids* dengan jumlah kluster yang sama terbentuk seperti pada Tabel 5. Hasil ini menjadi dasar dalam memilih konfigurasi kluster yang paling optimal untuk analisis lebih lanjut.

Tabel 5. *K-Means* dan *K-Medoids* K=3

No.	Kabupaten/ Kota	Cluster K-Means	Cluster K- Medoids	Jumlah_ Kejahatan	Jumlah_ Pendudu k	Tidak_ Bekerja	Resiko_ Kejahatan_ Penduduk	Rasio_ Kejahatan_ Tidak_ Beke rja
1	Cilacap	cluster_1	cluster_0	0,180	0,981	0,979	0,070	0,026
2	Banyumas	cluster_1	cluster_0	0,203	0,884	0,607	0,096	0,075
...
34	Kota Pekalongan	cluster_0	cluster_2	0,037	0,438	0,141	0,048	0,107
35	Tegal	cluster_1	cluster_0	0,065	0,789	0,771	0,023	0,003

Nilai *DBI* yang dimiliki oleh *K-Means* sebesar 0,082, yang lebih kecil dibandingkan dengan *K-Medoids* dengan jumlah kluster yang sama, yaitu 0,084. Maka algoritma *K-Means* dipilih menjadi metode pengelompokan yang digunakan dalam proses analisis, isi dari klusters yang terbentuk dari *K-Means* dengan K=3 dapat dilihat pada Tabel 6.

Tabel 6. *K-Means* K=3

No.	Kabupaten/ Kota	Cluster K-Means	Jumlah_ Kejahatan	Jumlah_ Penduduk	Tidak_ Bekerja	Resiko_ Kejahatan_ Penduduk	Rasio_ Kejahatan_ Tidak Bekerja
1	Purbalingga	cluster_0	0,099	0,449	0,312	0,111	0,086
2	Banjarnegara	cluster_0	0,008	0,46	0,356	0,014	0,009
3	Kebumen	cluster_0	0,043	0,65	0,413	0,022	0,024
4	Purworejo	cluster_0	0,153	0,32	0,155	0,237	0,255
5	Wonosobo	cluster_0	0,015	0,385	0,233	0,034	0,038
6	Magelang	cluster_0	0,13	0,614	0,359	0,097	0,094
7	Boyolali	cluster_0	0,051	0,483	0,234	0,054	0,074
8	Klaten	cluster_0	0,225	0,589	0,267	0,182	0,216
9	Sukoharjo	cluster_0	0,088	0,398	0,134	0,116	0,19
10	Wonogiri	cluster_0	0,04	0,462	0,097	0,047	0,156
11	Karanganyar	cluster_0	0,04	0,41	0,203	0,057	0,074
12	Sragen	cluster_0	0,133	0,433	0,179	0,152	0,2
13	Grobogan	cluster_0	0,139	0,702	0,343	0,086	0,106
14	Blora	cluster_0	0,038	0,381	0,141	0,062	0,108
15	Rembang	cluster_0	0,017	0,25	0,053	0,074	0,173
16	Pati	cluster_0	0,193	0,629	0,326	0,143	0,153
17	Kudus	cluster_0	0,062	0,366	0,121	0,095	0,163
18	Jepara	cluster_0	0,099	0,554	0,199	0,083	0,143
19	Demak	cluster_0	0,121	0,565	0,351	0,1	0,09
20	Semarang	cluster_0	0,057	0,478	0,238	0,06	0,078
21	Temanggung	cluster_0	0,016	0,331	0,068	0,047	0,144
22	Kendal	cluster_0	0,051	0,463	0,355	0,058	0,039
23	Batang	cluster_0	0,096	0,342	0,277	0,148	0,097
24	Pekalongan	cluster_0	0,091	0,438	0,141	0,105	0,185
25	Kota Magelang	cluster_0	0,011	0,614	0,359	0	0,011
26	Kota Surakarta	cluster_0	0,187	0,178	0,093	0,455	0,442
27	Kota Salatiga	cluster_0	0	0	0	0,295	0,283
28	Kota Pekalongan	cluster_0	0,037	0,438	0,141	0,048	0,107
29	Cilacap	cluster_1	0,18	0,981	0,979	0,07	0,026
30	Banyumas	cluster_1	0,203	0,884	0,607	0,096	0,075
31	Pemalang	cluster_1	0,08	0,718	0,549	0,041	0,027
32	Tegal	cluster_1	0,106	0,789	0,771	0,05	0,017
33	Brebes	cluster_1	0,089	1	1	0,02	0
34	Tegal	cluster_1	0,065	0,789	0,771	0,023	0,003
35	Kota Semarang	cluster_2	1	0,478	0,238	1	1

3.6. Deployment (Penerapan)

Pada tahap ini, hasil pengelompokan disampaikan kepada pihak yang berwenang atau pembuat kebijakan sebagai dasar dalam merumuskan strategi penanggulangan kejahatan yang lebih terarah dan efektif. Hasil evaluasi menunjukkan bahwa salah satu metode klasterisasi menghasilkan nilai indeks validitas yang lebih rendah dibandingkan metode lainnya pada jumlah klaster yang sama, yaitu 0,082 berbanding 0,084. Berdasarkan performa tersebut, metode dengan indeks yang lebih rendah dipilih sebagai dasar dalam proses pengelompokan data. Oleh karena itu, metode *K-Means* dengan performa terbaik dipilih sebagai dasar dalam proses analisis dan digunakan sebagai bahan pertimbangan utama dalam perumusan kebijakan pencegahan dan penanggulangan kejahatan di wilayah Provinsi Jawa Tengah. Berikut penamaan klaster yang dapat dilihat pada Tabel 7.

Tabel 7. Penamaan Klaster

Klaster	Nama Klaster	Deskripsi Singkat
Klaster 0	Wilayah Moderat dan Stabil	Risiko kejahatan sedang, wilayah beragam
Klaster 1	Wilayah Kepadatan Tinggi dengan Risiko Rendah	Penduduk tinggi, tapi risiko kejahatan sangat rendah
Klaster 2	Wilayah Risiko Tinggi	Risiko kejahatan sangat tinggi, perlu prioritas penanganan

4. KESIMPULAN

Penelitian ini menegaskan urgensi penerapan metode klastering dalam menganalisis pola kriminalitas di Provinsi Jawa Tengah. Melalui pendekatan algoritma *K-Means* dan *K-Medoids*, wilayah-wilayah dengan karakteristik sosial dan tingkat kriminalitas yang serupa berhasil dikelompokkan ke dalam klaster yang merepresentasikan kondisi lokal secara lebih objektif. Hasil clustering menunjukkan bahwa jumlah klaster optimal untuk kedua algoritma adalah 3 klaster, berdasarkan evaluasi menggunakan *DBI*. Nilai *DBI* terendah diperoleh oleh *K-Means* sebesar 0,082, sedikit lebih baik dibandingkan *K-Medoids* dengan nilai 0,084, yang menandakan bahwa pengelompokan menggunakan *K-Means* menghasilkan klaster yang lebih sama dan terbagi secara optimal. Dengan demikian, algoritma *K-Means* dengan 3 klaster dipilih sebagai model terbaik dalam pengelompokan data kriminalitas dan dijadikan dasar untuk pertimbangan kebijakan lebih lanjut. Hasil ini menunjukkan bahwa teknik clustering mampu mengidentifikasi pola tersembunyi dalam data kriminalitas, sehingga dapat digunakan sebagai dasar dalam perumusan strategi penanggulangan yang lebih tepat sasaran. Implikasi praktis dari penelitian ini mengarah pada pemanfaatan model berbasis data untuk mendukung perencanaan kebijakan oleh instansi pemerintah daerah dan aparat penegak hukum, khususnya dalam menyusun strategi pencegahan kejahatan yang kontekstual dan responsif terhadap karakteristik wilayah.

Meskipun demikian, penelitian ini memiliki sejumlah keterbatasan yang perlu dicermati. Variabel yang digunakan masih terbatas pada jumlah penduduk dan jumlah penduduk tidak bekerja, sehingga belum mencakup faktor-faktor sosial ekonomi lain yang juga berkontribusi terhadap tingkat kriminalitas, seperti tingkat pendidikan, tingkat kemiskinan, dan kepadatan penduduk. Selain itu, penggunaan algoritma klastering hanya terbatas pada *K-Means* dan *K-Medoids*, sehingga belum dapat menggambarkan kemungkinan hasil yang diperoleh melalui algoritma lainnya.

Untuk itu, penelitian selanjutnya disarankan untuk memperluas cakupan variabel yang dianalisis serta mengeksplorasi metode klastering alternatif seperti *DBSCAN* atau *Hierarchical Clustering* guna memperoleh sudut pandang yang lebih komprehensif terhadap struktur data. Secara keseluruhan, penelitian ini berkontribusi terhadap pengembangan pendekatan analitik berbasis data dalam pengelompokan wilayah rawan kejahatan. Kontribusi tersebut diharapkan dapat mendukung formulasi kebijakan publik yang lebih efektif, adaptif, dan berbasis bukti dalam rangka meningkatkan kualitas tata kelola keamanan dan ketertiban masyarakat secara berkelanjutan.

DAFTAR PUSTAKA

- [1] D. Yuliyanti and M. Martanto, "Clustering Tingkat Kejahatan Kriminal Menggunakan Metode K-Means Di Wilayah Kabupaten Cirebon," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 6, pp. 3509–3514, 2024, doi: 10.36040/jati.v7i6.8894.
- [2] B. P. S. P. J. Tengah, "Jumlah Kejahatan yang Dilaporkan, Risiko Penduduk Terkena Kejahatan per 100.000 Penduduk, Persentase Penyelesaian Kejahatan, dan Selang Waktu Terjadinya Kejahatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2023." <https://jateng.bps.go.id/>
- [3] N. Rohman and A. Wibowo, "Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall," *SINTECH (Science Inf. Technol. J.,* vol. 7, no. 1, pp. 49–58, 2024, doi: 10.31598/sintechjournal.v7i1.1507.

- [4] H. T. A. Simanjuntak, P. E. P. Silaban, J. K. S. Manurung, and V. H. Sormin, “Klasterisasi Berita Bahasa Indonesia Dengan Menggunakan K-Means Dan Word Embedding,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 3, pp. 641–652, 2023, doi: 10.25126/jtiik.20231026468.
- [5] A. Z. Kamalia and I. Nawangsih, “Identifikasi Pola Tingkat Kesenjangan Ketuntasan Pendidikan Di Indonesia Dengan Menggunakan Metode K-Medoids Cluster Ing Identification Of Patterns In Educational Completion Disparities In Indonesia Using The K-Medoids Clustering Method,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 2, pp. 321–330, 2025, doi: 10.25126/jtiik.2025129219.
- [6] A. Hoerunnisa, G. Dwilestari, F. Dikananda, H. Sunana, and D. Pratama, “Komparasi Algoritma K-Means Dan K-Medoids Dalam Analisis Pengelompokan Daerah Rawan Kriminalitas Di Indonesia,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 103–110, 2024, doi: 10.36040/jati.v8i1.8249.
- [7] G. Nurcahya, A. Wibowo, and D. Kristanto, “Perbandingan Kinerja Algoritma K-Means dan K-Medoids Dalam Klasterisasi Jumlah Tindak Pidana Kejahatan Berbasis Wilayah Kepolisian Daerah,” *SINTECH (Science Inf. Technol. J.)*, vol. 6, no. 3, pp. 162–172, 2023, doi: 10.31598/sintechjournal.v6i3.1457.
- [8] Z. Wahidah and D. T. Utari, “Implementation K-Means Algorithm to Group Provinces By Factors Influenced Criminal Act in Indonesia in 2019,” *Enthusiastic Int. J. Appl. Stat. Data Sci.*, vol. 2, no. 1, pp. 37–46, 2022, doi: 10.20885/enthusiastic.vol2.iss1.art5.
- [9] N. Azizah, A. Fauzi, T. Rohana, and S. Faisal, “Perbandingan Metode K-Means dan K-Medoids Untuk Clustering Jenis Kriminalitas,” *Buuld. Informatics, Technol. Sci.*, vol. 6, no. 2, pp. 1011–1019, 2024, doi: 10.47065/bits.v6i2.5723.
- [10] M. A. Z. Ramadhan, R. A. Saputra, and J. Nangi, “Klasterisasi Daerah Rawan Kriminalitas di Sulawesi Tenggara Menggunakan Metode K-Means Clustering,” *JOINTER J. Informatics Eng.*, vol. 5, no. 01, pp. 1–6, 2024, doi: 10.53682/jointer.v5i01.289.
- [11] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [12] J. Adiputra and D. Mahdiana, “Analisis Sentimen Dengan Algoritma Support Vector Machine Terhadap Penyakit Hepatitis Akut Misterius,” *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 1, pp. 1–8, 2023, doi: 10.36080/idealis.v6i1.2985.
- [13] W. Warisa and N. Nurahman, “Perbandingan Performa Cluster Model Algoritma K-Means Dalam Mengelompokkan Penerima Bantuan Program Keluarga Harapan,” *J. Sist. Info. Bisnis*, vol. 13, no. 1, pp. 20–28, 2023, doi: 10.21456/vol13iss1pp20-28.
- [14] S. Suraya, M. Sholeh, and D. Andayati, “Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa,” *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.
- [15] D. Haversyalapa, S. Puspasari, and R. Gustriansyah, “Klasterisasi Pixel Citra Koleksi Foto Museum Monpera Dengan Metode K-Means Pada Aplikasi Augmented Reality,” *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 2, pp. 189–199, 2024, doi: 10.36080/idealis.v7i2.3175.
- [16] Y. C. Fadilah, A. Sani, and A. Andrianingsih, “Applying K-Means Clustering For Grouping Papua ’ S Districts,” *JITK J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 10, no. 3, pp. 543–553, 2025, doi: 10.33480/jitk.v10i3.5865.APPLYING.
- [17] N. R. Saputra, G. Z. Muflih, and T. Informatika, “Pengelompokan Wilayah Indonesia Berdasarkan Komponen Indeks Pembangunan Manusia dengan Pendekatan Algoritma K-Means Clustering,” *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, pp. 156–167, 2025.
- [18] M. Wahyudi, D. Trianda, and L. Pujiastuti, “Implementation Of K-Medoids Method For Heart Disease Prediction Using Quantum Computing And Manhattan,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 3, pp. 469–478, 2025, doi: 10.33480/jitk.v10i3.5637.IMPLEMENTATION.
- [19] P. P. Allorerung, A. Erna, and M. Bagussahrir, “Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024.
- [20] H. Heru, S. Sunardi, and A. Amelia, “Identifikasi Kesalahan Siswa Dalam Menyelesaikan Soal Tipe Hots (Higher Order Thinking Skill) Materi Operasi Aljabar Di SMP Muhammadiyah 4 Palembang,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 23, pp. 656–664, 2023.