Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

Perbandingan Metode *K-Means* dan *K-Medoids* Untuk *Clustering* Jenis Kriminalitas

Nurul Azizah, Ahmad Fauzi, Tatang Rohana, Sutan Faisal*

Ilmu Komputer, Teknik Informatika, Universitas Buana Perjuangan Karawang, Indonesia
Email: if20.nurulazizah@mhs.ubpkarawang.ac.id, afauzi@ubpkarawang.ac.id, tatang.rohana@ubpkarawang.ac.id, sutan.faisal@ubpkarawang.ac.id
Email Penulis Korespondensi: sutan.faisal@ubpkarawang.ac.id
Submitted: **01/08/2024**; Accepted: **10/09/2024**; Published**: 12/09/2024**

Abstrak−Kriminalitas di Indonesia mencakup tindakan melanggar hukum, norma sosial, dan agama yang menyebabkan kerugian ekonomi dan psikologis serta ketegangan sosial dalam masyarakat. Kejahatan seperti pencurian, kekerasan, penipuan, dan narkoba sering dipicu oleh faktor-faktor seperti kemiskinan dan kondisi lingkungan yang mendukung perilaku kriminal. Penelitian ini perlu dilakukan untuk mengatasi masalah kriminalitas yang kompleks dan berdampak luas di Indonesia, khususnya di Kabupaten Karawang. Dengan meningkatnya kejahatan seperti pencurian, kekerasan, penipuan, dan narkoba, yang sering kali dipicu oleh faktor-faktor seperti kemiskinan dan kondisi lingkungan, diperlukan pendekatan yang lebih efektif untuk memahami dan menangani masalah ini. Penelitian ini menggunakan teknik data mining, khususnya analisis klaster, untuk mengelompokkan jenis-jenis kejahatan. Tujuannya adalah untuk mengidentifikasi pola-pola kriminalitas yang ada dan memahami faktor-faktor yang mempengaruhi penyebarannya. Dengan demikian, hasil penelitian ini dapat membantu pihak berwenang dalam mengembangkan strategi pencegahan dan penanganan kejahatan yang lebih tepat sasaran, sehingga dapat meminimalisir dampak negatif kriminalitas di wilayah tersebut. Selain itu, penelitian ini juga berkontribusi pada peningkatan pengetahuan mengenai metode yang paling efektif dalam analisis data kriminalitas, yang dapat diaplikasikan di daerah lain dengan permasalahan serupa. Hasil penelitian menunjukkan bahwa algoritma K-Means lebih efektif dibandingkan K-Medoids dalam menangani variabilitas data, dengan nilai Silhouette Coefficient sebesar 0.482 dan Davies Bouldin Index 0.915. Implementasi algoritma ini diharapkan dapat mempermudah identifikasi dan penanganan kejahatan di wilayah tersebut.

Kata Kunci: Kriminalitas; Data Mining; K-Means; K-Medoids; Clustering

Abstract−Crime in Indonesia includes acts that violate the law, social norms and religion which cause economic and psychological losses as well as social tensions in society. Crimes such as theft, violence, fraud and drugs are often triggered by factors such as poverty and environmental conditions that support criminal behavior. This research needs to be carried out to overcome the complex and far-reaching crime problem in Indonesia, especially in Karawang Regency. With crimes such as theft, violence, fraud and drugs on the rise, often fueled by factors such as poverty and environmental conditions, a more effective approach is needed to understand and address these problems. This research uses data mining techniques, especially cluster analysis, to group types of crime. The aim is to identify existing crime patterns and understand the factors that influence their spread. Thus, the results of this research can help the authorities in developing more targeted crime prevention and handling strategies, so as to minimize the negative impact of crime in the area. Apart from that, this research also contributes to increasing knowledge regarding the most effective methods for analyzing crime data, which can be applied in other areas with similar problems. The results of the research show that the K-Means algorithm is more effective than K-Medoids in handling data variability, with a Silhouette Coefficient value of 0.482 and a Davies Bouldin Index of 0.915. It is hoped that the implementation of this algorithm will make it easier to identify and handle crimes in the area.

Keywords: Crime; Data Mining; K-Means; K-Medoids; Clustering

1. PENDAHULUAN

Kejahatan adalah sebuah perbuatan dan perilaku yang melanggar hukum, norma sosial, dan norma agama yang berlaku di negara kesatuan Republik Indonesia sehingga menimbulkan kerugian ekonomi dan psikologis. Perilaku kriminal menyimpang adalah perilaku yang melanggar norma kehidupan sosial dan ketertiban sosial, dapat menimbulkan ketegangan individual dan ketegangan sosial dalam masyarakat, serta menimbulkan ancaman nyata atau potensial terhadap kelangsungan masyarakat[1].Kejahatan adalah tindakan negatif. Seringkali perbuatan tersebut menimbulkan kerugian bagi banyak pihak, dan individu yang melakukannya disebut penjahat. Ini melibatkan berbagai perbuatan yang dianggap melanggar hukum, seperti pencurian, kekerasan, penipuan, dan narkoba[2].

Seringkali, perilaku kriminal juga terjadi karena berbagai faktor kondisi, seperti meningkatnya kemiskinan, kondisi lingkungan yang mendukung kejahatan yang dilakukan seseorang, dan kebencian terhadap seseorang. Kejahatan merupakan isu yang hangat saat ini. Karena banyaknya ragam kejahatan, fenomena kriminalitas bermunculan tanpa henti di masyarakat[3] Kriminalitas yang sering terjadi seperti pembunuhan, perampokan, dan peristiwa yang lain dapat menimbulkan kekhawatiran di kalangan masyarakat [4]. Setiap warga negara seharusnya memiliki hak untuk merasa aman. Hal ini diatur dalam Pasal 28G ayat (1) Undang-Undang Dasar Negara Republik Indonesia Tahun 1945, yang menyatakan: "Setiap orang berhak atas perlindungan diri, keluarga, kehormatan, harkat, martabat, serta harta benda yang dimilikinya, dan berhak atas perlindungan dari rasa takut dalam melakukan atau tidak melakukan sesuatu yang dapat mengancam hak asasi manusia." Dapat dipahami bahwa kejahatan adalah tindakan yang dapat menimbulkan berbagai masalah dan ketidakstabilan dalam kehidupan masyarakat.[5].

Untuk meminimalisir terjadinya kriminalitas di Kabupaten Karawang, penting untuk melakukan pengelompokan jenis kriminalitas. Tujuannya adalah untuk mengidentifikasi pola-pola kriminalitas yang ada dan

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

memahami faktor-faktor yang mempengaruhi penyebarannya. Dengan demikian, hasil penelitian ini dapat membantu pihak berwenang dalam mengembangkan strategi pencegahan dan penanganan kejahatan yang lebih tepat sasaran, sehingga dapat meminimalisir dampak negatif kriminalitas di wilayah tersebut. Selain itu, penelitian ini juga berkontribusi pada peningkatan pengetahuan mengenai metode yang paling efektif dalam analisis data kriminalitas, yang dapat diaplikasikan di daerah lain dengan permasalahan serupa. Oleh karena itu, diperlukan pengelolaan data yang dapat menangani masalah tersebut, yaitu *clustering* jenis kriminalitas di kabupaten Karawang dengan menggunakan teknik data mining, khususnya analisis klaster *(clustering)*[6].

Clustering adalah salah satu subkategori dari data mining yang melibatkan proses pembagian sampel yang memiliki kesamaan ke dalam kelompok-kelompok yang disebut klaster. Setiap klaster berisi sampel-sampel yang memiliki kemiripan satu sama lain, tetapi berbeda dengan sampel-sampel yang terdapat dalam kelompok lain [7]. Dari berbagai metode dalam analisis klaster, terdapat dua jenis analisis klaster yang memiliki algoritma yang erat kaitannya, yakni K-Means dan K-Medoids clustering.[8]. Dalam penelitian ini, pengelompokan data dilakukan dengan menerapkan dua metode, yaitu Algoritma K-Means dan K-Medoids, untuk membandingkan kedua metode tersebut. Tujuannya adalah untuk memperoleh klaster yang paing efektif berdasarkan hasil perbandingan[9]. Metode K-Means yaitu mengelompokkan data kedalam beberapa klaster berdasarkan jarak[10]. Penerapan algoritma K-Means dan K-Medoids terbukti efektif dalam mengatasi permasalahan yang ada. Dengan Menerapkan algoritma K-Means dan K-Medoids terbukti efektif dalam mengatasi masalah yang ada. Dengan memanfaatkan kedua metode tersebut dalam penelitian ini, diharapkan proses identifikasi jenis kejahatan dengan tingkat kejahatan rendah dapat menjadi lebih mudah, sehingga membantu dalam menangani masalah yang ada[11].

Pada penelitian sebelumnya yang dilakukan oleh Hidayat R [12], Menerapkan Algoritma K-Means untuk mengelompokan daerah yang beresiko tinggi terhadap kejahatan di wilayah kabupaten Solok, berdasarkan data dari bulan januari dan desember 2021, maka di dapatkan hasil cluster tertinggi (CI) + = 2 kecamatan yaitu kecamatan kubung dan kecamatan sangir. Sedangkan untuk cluster rendah (C2) = 12 kecamatan, yaitu kecamatan bukit sundi, gunung talang, hiliran gumanti, jujuhan KPGD, Lembah gumanti, lembang jaya, lubuk sikarah, pauh duo, paying sekaki, sangir batang hari dan Sungai pagu. Mayona Y [13], menggunakan metode K-Means untuk mengelompokan Tingkat kejahatan Negeri Binjai dengan metode K-Means, Kluster 1 dengan centroid = 3, 1.33, 1.33 di kejaksaan negeri binjai, jenis kejahatan yang ditangani adalah Narkotika. Klaster 2 dengan centroid = 3, 7, 7 menunjukkan bahwa jenis kejahatan yang paling dominan adalah pencurian. Klaster 3 dengan centroid = 2.5, 7, 7 dengan jenis kejahatan adalah penipuan.

Pada penelitian yang dilakukan oleh Luchia N [14], dalam penelitian ini, perbandingan antara *K-Means* dan *K-Medoids* untuk pengelompokan data kemiskinan di Indonesia menunjukkan bahwa algoritma *K-Means* lebih unggul dibandingkan *K-Medoids*. Hal ini dibuktikan dengan nilai DBI terbaik untuk *K-Means*, yaitu 0.041, pada percobaan dengan K=8. Zuhdi Muzakkiy [15], juga melakukan penelitian tentang Penggunaan algoritma K-Means dalam metode clustering untuk menganalisis tindak kriminal melibatkan data dengan 9 atribut. Namun, dari 12.491 data yang ada, masih terdapat nilai yang hilang, sehingga perlu dilakukan imputasi menggunakan metode rata-rata atau frekuensi. Hasil pemodelan menunjukkan bahwa data lebih optimal jika dikelompokkan menjadi 4 klaster, karena skor Silhouette tertinggi diperoleh pada 4 klaster. Selain itu, salah satu dari 4 klaster yang diidentifikasi menjadi yang paling rentan terhadap kriminalitas

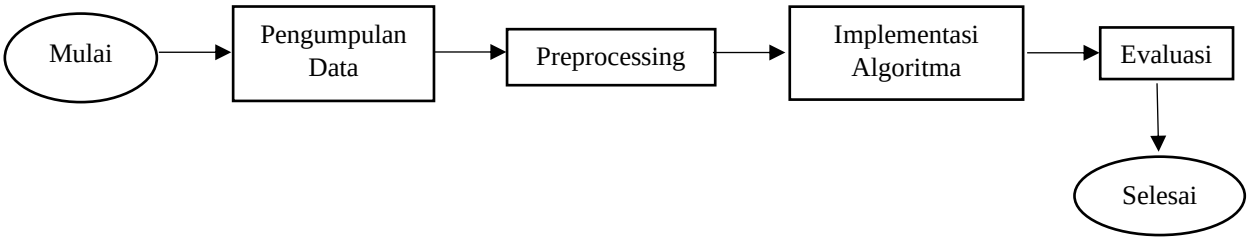
Abbas A [16] dalam penelitiannya dengan topik Implementasi clustering unsupervised learning menggunakan teknik pemetaan K-Means,Statistik yang digunakan meliputi 24 klasifikasi industri, dengan jumlah karyawan manufaktur di industri 2017-2019. Cluster Besar (E1) dan Cluster Rendah (E2). Parameter Davies Bouldin Index (DBI) dengan dbi nilai 0,929 digunakan untuk mengevaluasi cluster (k=2). Temuan menunjukkan lima manufaktur sektor-sektor cluster tinggi di 5 cluster dan 19 sektor manufaktur cluster kecil di 0 gugus. Untuk setiap cluster nilai centroidnya adalah 1,67; 1,64; 1,592 (klaster 1/E1) dan 0,348; 0,343;0,3447 (cluster 0/E2). Penelitian ini dilakukan oleh Dastriana dkk [17]Dalam penerapan algoritma K-Means dan K-Medoids untuk melakukan pengelompokan kabupaten dan kota di Provinsi Jawa Tengah berdasarkan produksi peternakan, hasil pengelompokan dari kedua metode menunjukkan bahwa metode yang paling optimal adalah K-Medoids. Pada k = 7, metode K-Medoids berhasil memberikan klaster yang optimal untuk data jumlah produksi ternak. Sedangkan AM Siregar [18] pada penelitian dengan tentang pengelompokan sektor pertumbuhan ekonomi Indonesia menggunakan algoritma K-Means menunjukkan bahwa hasil clustering dengan algoritma ini mengidentifikasi klaster rendah yang mencakup sektor-sektor seperti Pertanian, Kehutanan, dan Perikanan; Industri Pengolahan; Penyediaan Air, Pengelolaan Sampah, Limbah, dan Daur Ulang; Perdagangan Grosir dan Eceran; Reparasi Kendaraan Bermotor dan Sepeda Motor; Administrasi Pemerintahan; Pertahanan dan Jaminan Sosial Wajib; serta Jasa Lain-lain. Selama periode 6 tahun yang diuji, sektor-sektor ini menunjukkan peningkatan yang kurang memuaskan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam penelitian ini, terdapat beberapa tahapan penting yang harus dilalui untuk mencapai hasil yang diinginkan. Setiap tahapan memiliki peran yang spesifik dalam proses penelitian dan berkontribusi pada pemahaman keseluruhan tentang masalah yang diteliti. Pada Gambar 1 merupakan langkah-langkah penelitian yang akan dilakukan :

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723



Gambar 1. Tahapan Penelitian

Gambar 1 memberikan gambaran visual dari tahapan-tahapan penelitian yang dilakukan untuk mencapai hasil yang diinginkan. Setiap tahapan memiliki peran khusus dan saling terkait untuk memastikan bahwa proses penelitian dilakukan secara sistematis dan hasilnya akurat serta dapat dipercaya. Dengan mengikuti tahapan ini, peneliti dapat mengelola dan menganalisis data dengan cara yang efektif untuk mencapai tujuan penelitian. Pada Tabel 1, sumber data yang digunakan berasal dari Data Kriminalitas Polres Kabupaten Karawang. Data yang dianalisis mencangkup statistik kejahatan di kabupaten Karawang selama tahun 2021-2023.

Tabel 1. Sample Data

No	Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun
1	Pencurian dengan pemberatan	122	151	2021
2	Pencurian ranmor R-2	138	174	2021
3	Pencurian ranmor R-3	0	0	2021
4	Pencurian ranmor R-4	29	20	2021
5	Pencurian ranmor R-6	1	0	2021
…	…	…	…	…
329	UU Kesehatan	1	0	2023
330	UU HAKI	1	0	2023
331	UU BPJS	0	0	2023
332	UU Lingkungan hidup	0	o	2023
333	UU Perbankan	0	0	2023

Setelah mengumpulkan data sesuai dengan atribut yang akan digunakan, langkah selanjutnya adalah melakukan *pre-processing* data. Tujuannya adalah untuk menghilangkan duplikasi, menangani nilai yang hilang *(missing value)*, serta memperbaiki kesalahan-kesalahan yang mungkin ada pada data set baru dalam format excel. Setelah melalui proses *pre-processing*, data yang telah diperbaiki akan disimpan dalam sebuah data set baru menggunakan *Microsoft Office Excel.*

Setelah menyelesaikan tahap *preprocessing*, langkah berikutnya adalah menerapkan algoritma *K-Means* dan K-Medoids.

a. *K-Means*

Algoritma K-Means adalah metode *clustering* yang mengelompokkan data ke dalam beberapa klaster berdasarkan jarak[19]. *K-Means* termasuk dalam kategori algoritma unsupervised, yang artinya model ini tidak memerlukan label atau supervisi dalam data. Berikut adalah tahapan-tahapan yang dilibatkan dalam algoritma *K-Means* :

1. Menetapkan k sebagai jumlah klaster yang ingin dibentuk.
2. Ukur jarak antara setiap titik data dan pusat klaster.
3. Hitung jarak objek pada *centroid* menggunakan perhitungan Euclidean.

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad (1)$$

Keterangan :

D : Euclidean, yaitu jarak “lurus” anatar dua titik dalam ruang dua dimensi atau lebih. Jarak ini sering digunakan untuk menentukan kedekatan atau kesamaan antara objek dalam sebuah dataset.

I : Jumlah yang merujuk pada jumlah total objek yang di analisis atau dikelompokkan.

(x1,y1) : Untuk menentukan letak masing-masing objek dalam suatu bidang atau ruang yang di analisis

(x2,y2) : Merupakan titik acuan utama untuk menentukan seberapa dekat setiap objek dengan pusat klaster tersebut

4. Tentukan pusat klaster yang baru dengan menggunakan rumus.
5. Ulangi Langkah 2 hingga posisi data tetap tidak berubah

b. *K-Medoids*

Algoritma *K-Medoids* adalah metode *clustering* yang mirip dengan K-Means, tetapi dengan perbedaan mendasar dalam cara menentukan pusat klaster. Sementara K-Means menggunakan nilai rata-rata (mean) sebagai pusat klaster, K-Medoids menggunakan objek nyata sebagai medoid, yang berfungsi sebagai pusat klaster. K-Medoids

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

menggunakan medoid sebagai representasi klaster dan pusat perhitungan jarak. Berikut merupakan tahapan dalam algoritma K-Medoids:

1. Tetapkan pusat klaster sejumlah k (jumlah klaster).
2. Tempatkan setiap data (objek) ke klaster terdekat berdasarkan ukuran jarak Euclidean..
3. Pilih objek secara acak dalam tiap klaster sebagai calon medoid baru.
4. Ukur jarak setiap objek dalam klaster ke calon medoid baru.
5. Hitung total simpangan (S) dengan membandingkan nilai total jarak yang baru dengan nilai total jarak yang lama. Jika S < 0, tukar objek dengan data klaster untuk membuat medoid yang baru.
6. Ulangi langkah 3 hingga 5 sampai medoid tidak lagi berubah, sehingga kamu dapat menentukan klaster dan anggotanya masing-masing.

Hasil *clustering* menggunakan metode *K-Means* dan *K-Medoids* memiliki nilai yang belum pasti, sehingga untuk menilai kualitas *clustering* yang diperoleh perlu dilakukan perbandingan dengan hasil dari algoritma lain. Oleh karena itu, akan dilakukan evaluasi menggunakan metode *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI). Pada metode *Silhouette Coefficient* dilakukan evaluasi kualitas *clustering* berdasarkan hasil klaster yang diperoleh dengan menggabungkan aspek *cohesion* dan *separation*. Nilai yang dihasilkan oleh metode ini berkisar antara -1 hingga 1; nilai mendekati 1 menunjukkan bahwa kualitas pengelompokan klaster semakin baik[20]. Adapun *Davies-Bouldin Index* (DBI) adalah metode untuk mengevaluasi hasil *clustering*. Semakin kecil nilai DBI, semakin baik kualitas klaster tersebut.

3. HASIL DAN PEMBAHASAN

3.1 Hasil *Preprocessing*

Hasil *preprocessing* data dalam penelitian ini mencakup seleksi data, pembersihan data, transformasi data, dan normalisasi data. Berikut adalah tahapan-tahapan dari proses *preprocessing* data.

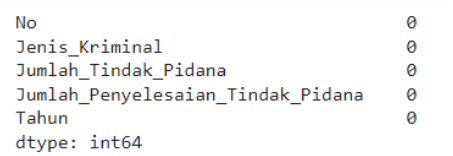
a. Seleksi Data
Tahap pertama *preprocessing* pada penelitian ini dimulai dari seleksi data, seleksi data digunakan untuk menghapus atribut yang tidak diperlukan karena dalam penelitian ini tidak semua atribut digunakan. Berikut adalah hasil seleksi data pada Tabel 2.

Tabel 2. *Seleksi Data*

Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun
Pencurian dengan pemberatan	122	151	2021
Pencurian ranmor R-2	138	174	2021
Pencurian ranmor R-3	0	0	2021
Pencurian ranmor R-4	29	20	2021
Pencurian ranmor R-6	1	0	2021
…	…	…	…
UU Kesehatan	1	0	2023
UU HAKI	1	0	2023
UU BPJS	0	0	2023
UU Lingkungan hidup	0	O	2023
UU Perbankan	0	0	2023

Pada Tabel 2 seleksi Data berisi informasi terkait berbagai jenis kriminalitas yang dikumpulkan dari dataset. Tabel ini menunjukkan rincian jumlah tindak pidana, jumlah penyelesaian tindak pidana, dan tahun untuk berbagai jenis kasus.

b. Data *Cleaning*
Tujuan dilakukannya data cleaning adalah untuk memastikan bahwa data yang digunakan dalam analisis atau pemodelan adalah akurat, konsisten, dan relevan. Data cleaning pada penelitian ini dilakukan untuk mengatasi nilai yang hilang, menghapus duplikasi data, dan mengeliminasi data yang tidak relevan. Proses data cleaning dapat dilihat pada Gambar 2.



Gambar 2. *Data Cleaning*

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

Gambar 2 menunjukkan bahwa dataset sudah bersih, tanpa adanya nilai yang hilang atau duplikasi data, sehingga proses selanjutnya dapat dilanjutkan.

c. Transformasi Data

Transformasi data bertujuan untuk mengubah data dari format teks ke format numerik, sehingga mempermudah proses pengolahan data. Proses transformasi data dapat dilihat pada Tabel 3..

Tabel 3. *Transformasi Data*

Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun
52	122	151	2021
45	138	174	2021
46	0	0	2021
47	29	20	2021
48	1	0	2021
…	…	…	…
92	1	0	2023
90	1	0	2023
89	0	0	2023
94	0	o	2023
106	0	0	2023

Pada Tabel 3 transformasi Data menyajikan data dari 5 data teratas dan 5 data terbawah yang telah mengalami proses transformasi dari bentuk aslinya. Transformasi data ini bertujuan untuk menyederhanakan dan menstandarkan format data sehingga lebih mudah dianalisis

d. Normalisasi Data

Normalisasi Dilakukan untuk mengubah nilai variabel sehingga dapat diukur pada skala yang umum, biasanya berada dalam rentang 0 hingga 1. Pada penelitian ini normalisasi dilakukan dengan menggunakan min-max. Berikut normalisasi data pada Tabel 4.

Tabel 4. Normalisasi Data

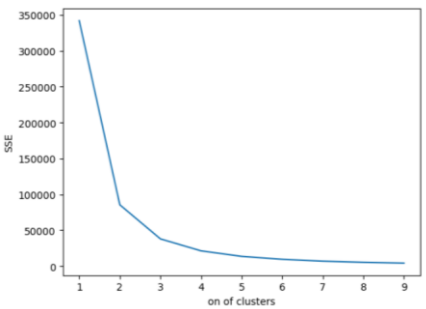
Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun
52	0.764706	0.843137	0.0
45	0.803922	0.882353	0.0
46	0.000000	0.000000	0.0
47	0.392157	0.352941	0.0
48	0.019608	0.352941	0.0

Tabel 4 merupakan hasil dari normalisasi data dari 5 data teratas yang berisi informasi tentang "Jenis Kriminal", "Jumlah Tindak Pidana", "Jumlah Penyelesaian Tindak Pidana", dan "Tahun". Normalisasi adalah proses untuk mengubah nilai-nilai dalam dataset ke dalam skala tertentu, biasanya antara 0 dan 1. Proses ini sangat penting dalam algoritma machine learning seperti K-Means agar fitur-fitur dengan skala berbeda dapat diperlakukan secara seimbang.

3.2 Implementasi Algoritma

Jumlah klaster dalam penelitian ini ditentukan dengan menggunakan Metode Elbow, yang berfungsi untuk menentukan nilai k optimal dalam Algoritma K-Means. Penetapan nilai k dilakukan dengan menganalisis nilai SSE (Sum of Squared Errors) yang ditampilkan dalam grafik pada Gambar 3.

a. Hasil Implementasi Algoritma *K-Means*



Gambar 3. *Grafik Metode Elbow*

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

Gambar 3 menunjukkan bahwa grafik mengalami perubahan yang signifikan, peningkatan grafik menjadi lebih lambat atau stabil, menunjukkan grafik yang lebih datar pada nilai k = 2. Berdasarkan nilai k tersebut, jenis kriminalitas dapat dikelompokkan ke dalam 2 klaster. Setelah menentukan jumlah klaster = 2, langkah berikutnya adalah menetapkan pusat klaster, yang dapat dilihat dalam Tabel 5.

Tabel 5. *Pusat Centroid Algoritma K-Means*

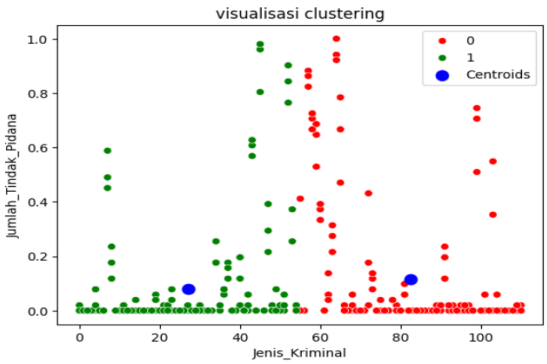
No.	Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun
1.	2.70000000	8.05704100	9.48306595	5.00000000
2.	8.25000000	1.14028945	1.06442577	5.00000000

Setelah pusat centroid ditetapkan, langkah berikutnya adalah mengelompokkan data yang telah dinormalisasi, kemudian menambahkan kolom "cluster" untuk menunjukkan hasil klastering. Hasil pengelompokan menggunakan Algoritma K-Means dapat ditemukan pada Tabel 6.

Tabel 6. *Hasil Custering Algoritma K-Means*

Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun	Cluster
52	0.764706	0.843137	0.0	0
45	0.803922	0.882353	0.0	0
46	0.000000	0.000000	0.0	0
47	0.392157	0.352941	0.0	0
48	0.019608	0.352941	0.0	0
…	…	…	…	…
92	0.019608	0.000000	1.0	1
90	0.019608	0.000000	1.0	1
89	0.000000	0.000000	1.0	1
94	0.000000	0.000000	1.0	1
106	0.000000	0.000000	1.0	1

Tabel 6 menunjukkan jumlah anggota di setiap klaster. Klaster 1 memiliki 177 kasus kriminal dan klaster 2 memiliki 156 kasus kriminal. Hasil visualisasi dari *clustering* dapat ditemukan pada Gambar 4.



Gambar 4. *Visualisasi Hasil Clastering*

Gambar 4 menunjukkan visualisasi hasil pengelompokan jenis kriminal menggunakan Algoritma K-Means, yang membagi data menjadi dua kelompok. Kelompok 1 diberi warna merah, sementara kelompok 2 diberi warna hijau. Titik centroid setiap kelompok diberi tanda dengan warna biru.

b. Hasil Implementasi Algoritma *K-Medoids*

Penerapan Algoritma K-Medoids untuk menentukan jumlah klaster = 2 dilakukan dengan Metode Elbow. Setelah jumlah klaster ditentukan, langkah berikutnya adalah menetapkan pusat centroid. Hasil penetapan pusat centroid menggunakan Algoritma K-Medoids ditampilkan pada Tabel 7.

Tabel 7. Pusat Centroid Algoritma K-Medoids

No.	Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun
1.	5.50000000	4.11764706	1.96078431	5.00000000
2.	1.90000000	0.00000000	0.00000000	5.00000000
3	9.20000000	1.96078431	0.00000000	5.00000000

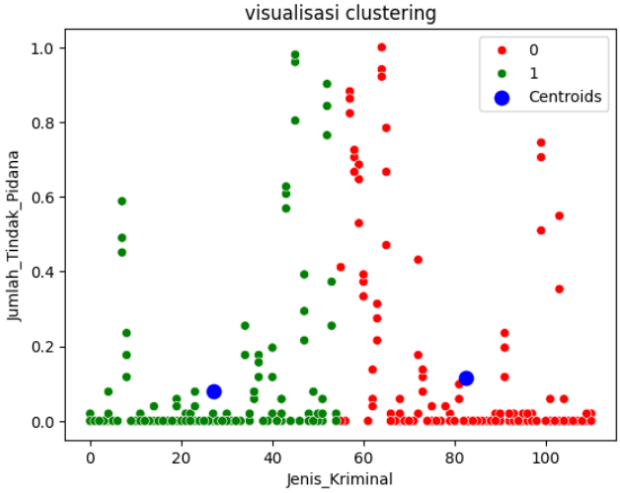
Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

Setelah pusat centroid ditetapkan, langkah berikutnya adalah mengelompokkan data yang telah dinormalisasi, dan menambahkan kolom "*cluster*" untuk menunjukkan hasil klastering. Hasil klastering menggunakan Algoritma *K-Medoids* dapat ditemukan pada Tabel 8.

Tabel 8. *hasil clustering Algoritma K-Medoids*

Jenis Kriminal	Jumlah Tindak Pidana	Jumlah Penyelesaian Tindak Pidana	Tahun	Cluster
52	0.764706	0.843137	0.0	0
45	0.803922	0.882353	0.0	0
46	0.000000	0.000000	0.0	0
47	0.392157	0.352941	0.0	0
48	0.019608	0.352941	0.0	0
…	…	…		…
92	0.019608	0.019608	1.0	2
90	0.019608	0.019608	1.0	2
89	0.000000	0.000000	1.0	2
94	0.000000	0.000000	1.0	2
106	0.000000	0.000000	1.0	2

Hasil dari pengelompokan menggunakan Algoritma K-Medoids yang tercantum dalam Tabel 8 menunjukkan jumlah anggota di setiap klaster. Klaster 1 memiliki 174 kasus kriminal, klaster 2 memiliki 159 kasus kriminal. Visualisasi hasil dari pengelompokan dapat ditampilkan dalam Gambar 5.



Gambar 5. *Visualisasi Hasil Klastering*

Gambar 5 menampilkan visualisasi hasil pengelompokan jenis kriminal menggunakan Algoritma K-Medoids, yang membagi data menjadi dua kelompok. Kelompok 1 ditandai dengan warna merah, kelompok 2 dengan warna hijau. Setiap kelompok memiliki titik centroid yang diberi dengan warna biru.

3.3 Evaluasi

Berikut adalah hasil evaluasi klaster menggunakan nilai *Silhoutte Coefficient*. *Silhouette Coefficient* digunakan untuk menilai seberapa baik klasterisasi dilakukan dan seberapa jelas klaster yang dihasilkan. Nilai yang lebih tinggi menunjukkan kualitas klasterisasi yang lebih baik. Dapat dilihat pada Tabel 9.

Tabel 9. Evaluasi Sillhoutte Coefficient

Algoritma	Klaster 1 (Tinggi)	Klaster 2 (Rendah)	Evaluasi *Sillhoutte Coefficient*
K-Means	177	156	0.48246138301771496
K-Medoids	174	159	0.4768601432018354

Tabel 9. Menampilkan hasil perbandingan evaluasi *cluster* antara Algoritma *K-Means* dan *K-Medoids* menggunakan Evaluasi *Silhouette Coefficient*. Evaluasi ini menunjukkan bahwa klasterisasi dengan Algoritma K-Means memperoleh hasil terbaik karena *K-Means* biasanya lebih efektif dalam menangani data dengan variabilitas yang tinggi karena menggunakan *centroid* sebagai representasi dari setiap *cluster*. *Centroid* merupakan titik tengah dari data dalam sebuah *cluster*, sehingga dapat menyesuaikan lebih baik terhadap variasi yang ada dalam data , Adapun hasil klaster terbaik berdasarkan *Silhouette Coefficient* sebesar 0.482. Berikut adalah hasil evaluasi klaster menggunakan nilai *Davies Bouldin Index* pada Tabel 10.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

Tabel 10. *Evaluasi Davies Bouldin Index*

Algoritma	Klaster 1 (Tinggi)	Klaster 2 (Rendah)	Evaluasi *Davies Bouldin Index*
K-Means	177	156	0.9156711502581246
K-Medoids	174	159	0.9741143987228782

Tabel 10 menampilkan hasil perbandingan evaluasi *cluster* antara Algoritma *K-Means* dan *K-Medoids* menggunakan Evaluasi *Davies Bouldin Index*. . Evaluasi ini menunjukkan bahwa klasterisasi dengan Algoritma *K-Means* memperoleh hasil terbaik karena *K-Means* mengoptimalkan fungsi tujuan yang secara langsung meminimalkan jarak kuadrat antara data dengan *centroid* terdekat. Ini membuat *K-Means* cenderung menghasilkan *cluster* yang lebih kompak dan terpisah, sehingga menghasilkan nilai *Davies Bouldin Index* (DBI) yang lebih rendah. Adapun hasil klaster terbaik berdasarkan *Davies Bouldin Index* adalah 0.915. Berdasarkan tabel 10 klasterisasi antara Algoritma K-Means dan K-Medoids menunjukkan bahwa K-Means memberikan hasil terbaik berdasarkan dua metrik evaluasi, yaitu Silhouette Coefficient dan Davies Bouldin Index. Hasil terbaik untuk Silhouette Coefficient diperoleh oleh K-Means dengan nilai 0.482, yang menunjukkan bahwa K-Means lebih efektif dalam menangani data dengan variabilitas tinggi. Hasil terbaik untuk Davies Bouldin Index juga diperoleh oleh K-Means dengan nilai 0.915, yang mengindikasikan bahwa K-Means menghasilkan klaster yang lebih kompak dan terpisah.

4. KESIMPULAN

Kejahatan adalah tindakan melanggar hukum, norma sosial, dan norma agama yang berlaku di Indonesia, yang sering kali menimbulkan kerugian ekonomi dan psikologis. Faktor-faktor seperti kemiskinan, kondisi lingkungan yang mendukung kejahatan, dan kebencian terhadap seseorang dapat menjadi pemicu terjadinya tindakan kriminal. Penelitian ini menekankan pentingnya mengidentifikasi pola-pola kriminalitas di Kabupaten Karawang dengan menggunakan teknik data mining, khususnya analisis klaster *(clustering)*, untuk membantu pihak berwenang dalam mengembangkan strategi pencegahan dan penanganan kejahatan. Dalam penelitian ini, dua metode *clustering*, yaitu *K-Means* dan *K-Medoids*, digunakan untuk mengelompokkan data kriminalitas. Hasil evaluasi menunjukkan bahwa Algoritma *K-Means* memberikan hasil yang lebih baik dibandingkan *K-Medoids* dalam hal kualitas *clustering*, ditunjukkan oleh nilai *Silhouette Coefficient* sebesar 0.482 dan *Davies Bouldin Index* 0.915. Dengan demikian, penggunaan Algoritma *K-Means* dianggap lebih efektif dalam mengidentifikasi jenis kejahatan dan dapat membantu dalam penanganan masalah kriminalitas di Kabupaten

REFERENCES

[1] J. FButarbutar, N. Budi Nugroho, dan W. Rista Maya, "Implementasi Data Mining untuk Mengelompokkan Daerah Rawan Tindakan Kriminal di Kota Medan Menggunakan Metode K-Means," *Jurnal CyberTech*, vol. 4, no. 3, 2021,doi :https://doi.org/10.53513/jct.v4i3.3793

[2] P. S. Rosiana *dkk.*, "Visualisasi Data Tindak Kejahatan Berdasarkan Jenis Kriminalitas di Kabupaten Karawang dengan Menggunakan Algoritma Clustering K-Means," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, hlm. 2830–7062, 2023, doi: 10.23960/jitet.v11i3%20s1.3347.

[3] R. Astuti dan F. Muhammad Basysyar, "Penerapan Data Mining Clustering Menggunakan Metode K-Means pada Data Tindak Kriminalitas di Polres Kabupaten Kuningan," Vol.8 No. 2, 2024. doi : https://doi.org/10.36040/jati.v8i2.8790

[4] D. Gultom *dkk.*, "Penerapan Algoritma K-Means untuk Mengetahui Tingkat Tindak Kejahatan di Daerah Pematangsiantar," *Jurnal Teknologi Informasi*, vol. 4, no. 1, pp 2580-7927, 2020.

[5] R. N. Fahmi, M. Jajuli, N. Sulistiyowati, "Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang Menggunakan Algoritma K-Means," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 4, no. 1, 2021, doi : https://doi.org/10.31539/intecoms.v4i1.2413

[6] H. S. Firdaus, A. L. Nugraha, B. Sasmito, M. Awaluddin, dan C. A. Nanda, "Perbandingan Metode Fuzzy C-Means dan K-Means untuk Pemetaan Daerah Rawan Kriminalitas di Kota Semarang," Vol. 4, No. 01, 2021. doi : https://doi.org/10.14710/halal.v%25vi%25i.9219

[7] Sekar Setyaningtyas, B. Indarmawan Nugroho, dan Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 2, hlm. 52–61, Okt 2022, doi: 10.21063/jtif.2022.v10.2.52-61.

[8] H. Dame Tampubolon, M. Safii, dan D. Suhendro, "Penerapan Algoritma K-Means dan K-Medoids Clustering untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi," vol. 2, no. 2, hlm. 6–12, 2021

[9] N. A. Kamilah, T. Rohana, dan A. Fauzi, "JURNAL MEDIA INFORMATIKA BUDIDARMA Implementasi Algoritma K-Means dan K-Medoids Dalam Klasterisasi Kasus Kekerasan Terhadap Perempuan," 2024, doi: 10.30865/mib.v8i2.7558.

[10] B. Biantara, T. Rohana, dan A. Ratna Juwita, "Perbandingan Algoritma K-Means dan DBSCAN untuk Pengelompokan Data Penyebaran Covid-19 Seluruh Kecamatan di Provinsi Jawa Barat," vol. IV, no. 1, pp 88-94, 2023.

[11] L. Adeliana, A. Mutoi Siregar, dan D. Sulistya Kusumaningrum, "Pengelompokan Kabupaten dan Kota di Indonesia Berdasarkan Hasil Produksi Daging Sapi Menggunakan Algoritma K-Means dan K-Medoids," vol. II, no. 1, 2021.

[12] R. Hidayat, "Clustering Menggunakan Algoritma K-Means untuk Mengelompokan Wilayah Rawan Kejahatan di Wilayah Kabupaten Solok." Vol.4 No.5, 2022. doi : https://doi.org/10.32639/jimmba.v4i5.169

Building of Informatics, Technology and Science (BITS)
Volume 6, No 2, September 2024, Page: 1011−1019
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i2.5723

[13] Y. Mayona, R. Buaton, dan M. Simanjutak, "Data Mining Clustering Tingkat Kejahatan Dengan Metode Algoritma K-Means (Studi Kasus : Kejaksaan Negeri Binjai)," *Agustus*, vol. 6, no. 3, 2022.
[14] J. Homepage *dkk.*, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," *Institut Riset dan Publikasi Indonesia*. Vol. 2, no. 2, hlm. 35–41, 2022. doi : https://doi.org/10.57152/malcom.v2i2.422
[15] I. Zuhdi Muzakkiy -, K. Husein -, K. Antonius -, K. Raihan Hidayat -, E. Emir Di Haryanto -, dan I. Paryudi -, "Penggunaan Algoritma K-Means Pada Metode Clustering Untuk Menganalisa Tindak Kriminal."Jurnal of Informatics and Advanced Computing. Vol.4 No.1, pp 16-21, 2023.
[16] A. Abbas *dkk.*, "Implementation of clustering unsupervised learning using K-Means mapping techniques," *IOP Conf Ser Mater Sci Eng*, vol. 1088, no. 1, hlm. 012004, Feb 2021, doi: 10.1088/1757-899x/1088/1/012004.
[17] M. Khoncita Dasriana Bau, Y. Setyawan, M. Titah Jatipaningrum, J. Statistika, F." Perbandingan Metode Algoritma K-Means dan K-Medoids pada Pengelompokan Kabupaten/Kota di Provinsi Nusa Tenggara Timur Berdasarkan Dimensi Indeks Pembangunan Manusia Tahun 2020," *"Jurnal Statistika Industri dan Komputasi"* vol. 08, no. 1, hlm. 48–57, 2023
[18] A. M. Siregar, "Pengelompokan Bidang Laju Pertumbuhan Ekonomi Indonesia Menggunakan Algoritma K-Means."*Accounting Information System*. Vol.2 No.2, 2019. doi: https://doi.org/10.32627/aims.v2i2.342
[19] A. B. Sarana, I. Yogyakarta, N. Sari, H. H. Handayani, dan A. M. Siregar, "Implementasi Clustering Data Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means," *Bianglala Informatika : Jurnal Komputer dan Informatika* vol. 11. No.1, 2023. doi : https://doi.org/10.31294/bi.v11i1.14762
[20] C. Ayudia, S. Fastaf, dan Y. Yamasari, "Analisa Pemetaan Kriminalitas Kabupaten Bangkalan Menggunakan Metode K-Means dan K-Means++," *Journal of Informatics and Computer Science*, vol. 03, 2022. doi: https://doi.org/10.26740/jinacs.v3n04.p534-546