

KLASTERISASI TINGKAT KRIMINALITAS DI INDONESIA MENGUNAKAN METODE K-MEANS

Shiva Divanti Natasya, Ahmad Faisol, Deddy Rudhistiar

Teknik Informatika, Institut Teknologi Nasional Malang

Jalan Raya Karanglo km 2 Malang, Indonesia

2118089@scholar.itn.ac.id

ABSTRAK

Tingginya tingkat kriminalitas di Indonesia menjadi permasalahan serius yang berdampak pada keamanan dan ketertiban masyarakat. Variasi jumlah dan jenis kejahatan di tiap provinsi, seperti tingginya kasus pencurian di Sumatera Utara dan pembunuhan di Jawa Timur, menunjukkan adanya ketimpangan kondisi sosial ekonomi antar wilayah. Sayangnya, informasi mengenai tingkat kriminalitas sering kali hanya diketahui secara reaktif setelah kasus terjadi, sehingga langkah pencegahan menjadi kurang optimal. Penelitian ini memiliki tujuan dalam rangka membangun sistem klasterisasi tingkat kriminalitas di Indonesia berdasarkan data sosial ekonomi dan jenis kejahatan yang dilaporkan. Metode yang dipergunakan yaitu K-Means Clustering dengan variabel berupa data pembunuhan, pencurian, narkoba, perkosaan, pencurian kendaraan bermotor, pengangguran, kemiskinan, dan IPM. Hasil penelitian menunjukkan bahwa sistem berhasil mengelompokkan provinsi dengan cukup baik, ditunjukkan oleh nilai *Silhouette Score* rata-rata sebesar 0,34, dengan nilai tertinggi terdapat pada Cluster 3 sebesar 0,38. Penelitian ini diharapkan dapat memberikan gambaran pola kriminalitas sebagai dasar pengambilan kebijakan yang lebih tepat sasaran. Untuk pengembangan selanjutnya, disarankan agar data mencakup seluruh kota di tiap provinsi dan metode K-Means++ digunakan untuk peningkatan akurasi klasterisasi.

Kata kunci : kriminalitas, klasterisasi, K-Means, data sosial ekonomi, *Silhouette Score*

1. PENDAHULUAN

Kriminalitas merupakan tindakan yang melanggar hukum dan dapat dikenai sanksi pidana sesuai ketentuan perundang-undangan [1]. Selain berdampak pada korban dan pelaku, kriminalitas juga menimbulkan rasa tidak aman di masyarakat, sehingga pemahaman terhadap pola dan tingkat kriminalitas di berbagai wilayah sangat diperlukan.

Sayangnya, informasi kriminalitas umumnya baru diketahui setelah kejadian terjadi, membuat upaya pencegahan menjadi reaktif. Diperlukan pendekatan yang lebih proaktif agar masyarakat dan pemangku kebijakan dapat mengambil langkah yang tepat.

Tingkat kriminalitas di Indonesia bervariasi antar provinsi. Sumatera Utara mencatat 8.837 kasus pencurian dan 5.306 kasus narkoba, sementara Jawa Timur mencatat 100 kasus pembunuhan. Sebaliknya, Kepulauan Bangka Belitung hanya mencatat 3 kasus pembunuhan [2]. Provinsi seperti Sulawesi Barat dan Kalimantan Utara bahkan sempat belum memiliki data tersendiri karena belum terbentuknya Polda secara mandiri [3].

Perbedaan ini dipengaruhi oleh berbagai faktor sosial-ekonomi seperti jumlah penduduk, pengangguran, kemiskinan, dan Indeks Pembangunan Manusia (IPM). Penelitian ini bertujuan mengkategorikan provinsi di Indonesia menurut tingkat kriminalitas dan kondisi sosial-ekonominya menggunakan metode K-Means Clustering, yaitu algoritma klasterisasi yang mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik. Diharapkan hasil klasterisasi dapat

menjadi dasar perumusan kebijakan yang lebih efektif dan tepat sasaran.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Pada penelitian berjudul “Klasterisasi Daerah Rawan Kriminalitas di Sulawesi Tenggara Menggunakan Metode K-Means Clustering”, penelitian ini bertujuan untuk mengkategorikan daerah rawan kriminalitas di Sulawesi Tenggara. Kriteria yang digunakan adalah jumlah kasus kejahatan dan persentase korban di setiap daerah. Hasil klasterisasi menunjukkan tiga kategori tingkat kriminalitas: rendah, sedang, dan tinggi. Evaluasi menggunakan Davies-Bouldin Index menghasilkan nilai 0,5904 [1].

Kemudian, pada penelitian yang berjudul “Clustering Tingkat Kejahatan Kriminal Menggunakan Metode K-Means di Wilayah Kabupaten Cirebon” dilakukan klasifikasi tingkat kriminalitas menggunakan data 123 kasus melalui algoritma K-Means. Hasil klasterisasi menunjukkan sembilan klaster yang terbagi ke dalam tiga tingkat kriminalitas. Nilai Davies-Bouldin Index terbaik diperoleh pada K=9 dengan nilai 0,122 [4].

Lalu, dalam penelitian yang berjudul “Sistem Informasi Geografis Pengelompokan Tingkat Kriminalitas Kota Malang Menggunakan Metode K-Means” metode K-Means digunakan untuk mengelompokkan penerima BST berdasarkan wilayah dan jumlah penerima. Penelitian ini bertujuan meningkatkan efektivitas distribusi bantuan [5].

Berikutnya, dalam penelitian berjudul “Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Penerima Bantuan Sosial Tunai

(BST) di Jawa Barat” mengelompokkan wilayah penerima BST menggunakan metode K-Means dengan tujuan meningkatkan efektivitas distribusi bantuan berdasarkan lokasi dan jumlah penerima [6]

Selanjutnya, dalam penelitian yang berjudul “Penerapan Algoritma K-Means untuk mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar” mengelompokkan enam kecamatan di Kota Pematangsiantar menjadi dua kluster tingkat kejahatan. Hasil menunjukkan bahwa Kecamatan Siantar Martoba memiliki kluster kejahatan tinggi, sedangkan Kecamatan Siantar Selatan memiliki kluster kejahatan rendah [7].

2.2. Data Mining

Data mining ialah suatu teknik yang dimanfaatkan dalam mengidentifikasi pola serta informasi tersembunyi dalam kumpulan data berukuran besar [8]. Proses ini juga dikenal sebagai penemuan pengetahuan (knowledge discovery), yang bertujuan menggali wawasan baru dari data yang sudah tersedia [9]. Satu dari beberapa metode yang umum dipergunakan pada data mining yaitu K-Means Clustering, yakni teknik pengelompokan data yang tidak mempunyai struktur hierarki. Metode ini mengelompokkan data ke dalam sejumlah kelompok menurut kesamaan karakteristik, yang mana dapat menunjang proses analisis dan mengambil keputusan.

2.3. Klasterisasi

Klasterisasi ialah metode dalam analisis data yang dipergunakan pada pengelompokan data ke dalam sejumlah kategori menurut tingkat kemiripan antar data. Proses ini melibatkan penilaian kesamaan antar data dengan memanfaatkan berbagai teknik pengukuran, seperti jarak Euclidean, *Manhattan*, atau *Minkowski*. Akurasi dan kualitas hasil klasterisasi sangat dipengaruhi oleh metode yang digunakan, karena pemilihan metode akan berdampak langsung terhadap hasil akhir pengelompokan [9].

2.4. K-Means Clustering

K-Means Clustering adalah satu dari beberapa metode unsupervised learning yang digunakan untuk membagi data ke dalam sejumlah k kelompok menurut kemiripan karakteristik, tanpa memerlukan label data sebelumnya [10]. Proses algoritmanya diawali dengan pemilihan pusat *cluster* secara acak, kemudian data diklasifikasikan ke dalam *cluster* berdasarkan jarak terdekat, dan langkah ini diulang hingga tercapai kondisi stabil [11]. Sebelum proses pengelompokan dilakukan, data terlebih dahulu dinormalisasi agar setiap variabel berada dalam skala yang sama, dengan demikian hasil klasterisasi dapat lebih akurat dan tidak bias terhadap variabel dengan nilai yang lebih besar.

Prosedur untuk mengorganisir data dengan Algoritma K-Means terdiri dari beberapa langkah sebagai berikut :

- Penentuan nilai k berdasarkan kebutuhan analisis data.

- Menentukan awal titik pusat kluster (*centroid*).
- Mengukur jarak dari setiap data terhadap setiap centroid dapat dilakukan dengan rumus Jarak Euclidean. Berikut adalah rumus Jarak Euclidean tersebut[1]:

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

Keterangan :

d = Jarak antara x dan y

x = Data

y = *Centroid*

i = Iterasi

n = Jumlah dari data

- Kelompokkan setiap data ke dalam kluster berdasarkan jarak minimum dengan *centroid*.
- Mendapatkan lokasi centroid baru dengan memperhitungkan rata-rata dari semua hasil untuk setiap centroid atau kluster.
- Lakukan langkah 3 hingga 5 berulang kali sampai kluster tidak berubah lagi.

2.5. Silhouette Score

Silhouette Score merupakan metode evaluasi yang dipergunakan dalam mengukur seberapa baik data telah dikelompokkan dalam proses klasterisasi. Penilaian ini didasarkan pada dua faktor utama: kedekatan data dalam satu kluster (kohesi) dan jaraknya terhadap kluster lain (pemisahan). Semakin tinggi nilai *Silhouette Score*, semakin baik kualitas pengelompokan yang dihasilkan [12]. Adapun rumus *sillhouette score* adalah sebagai berikut :

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Keterangan :

$S(i)$ = Nilai *Silhouette Score*

$a(i)$ = adalah nilai *range* jarak antara suatu titik data ke- i dengan seluruh titik lain yang ada pada kluster yang sama.

$b(i)$ = adalah nilai *range* jarak antara titik data ke- i dengan semua titik yang ada di kluster berbeda yang paling dekat dengannya.

Rata-rata nilai *Silhouette Score* memberikan informasi tentang seberapa baik suatu objek terintegrasi dalam kluster. Nilai ini berkisar dari -1 sampai 1. Semakin mendekati angka 1, semakin baik pembagian antara kluster, yang berarti objek berada lebih dekat dengan pusat kluster.

2.6. Database

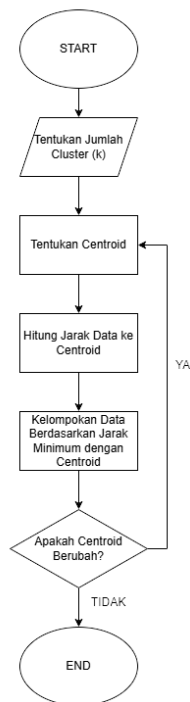
Basis data merupakan himpunan data yang disusun secara terstruktur pada sistem komputer, sehingga dapat diakses dan dikelola dengan mudah oleh berbagai program. Tujuan inti dari sistem basis data yaitu guna membentuk representasi abstrak dari data serta mempermudah interaksi antara pengguna dan sistem [13]. Di sisi lain, basis data juga dapat dipahami sebagai himpunan informasi yang tertata rapi, tersimpan di satu tempat, dan didesain guna memenuhi kebutuhan informasi pada suatu organisasi [14].

2.7. Laravel

Laravel merupakan kerangka kerja berbasis PHP yang banyak digunakan dalam pengembangan web karena kemudahan penggunaan dan fungsionalitasnya [15]. Framework ini mengadopsi arsitektur Model-View-Controller (MVC) dan membantu mempercepat proses pengembangan serta memudahkan pemeliharaan. Laravel bersifat sumber terbuka dengan lisensi MIT dan didistribusikan melalui platform GitHub [16].

3. METODE PENELITIAN

3.1. Flowchart Metode K-Means Clustering



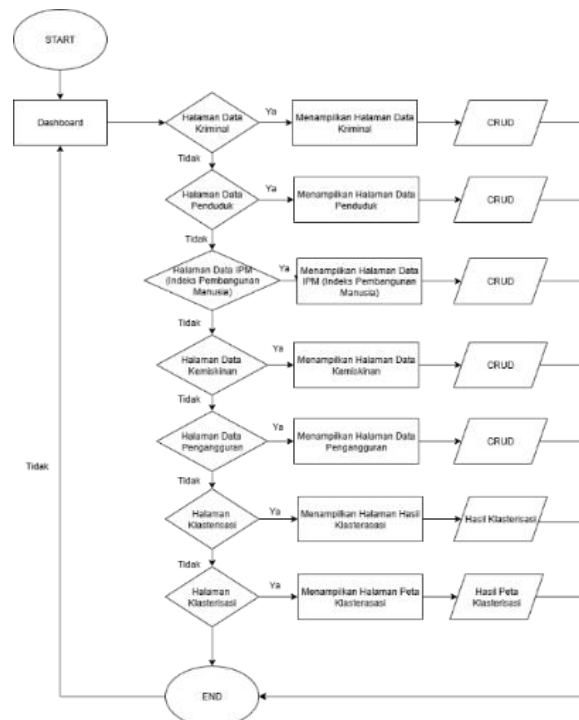
Gambar 1. Flowchart Metode K-Means Clustering

Gambar 1 menggambarkan proses flowchart dari metode K-Means. Langkah pertama dalam proses ini adalah menentukan jumlah kluster yang diperlukan. Selanjutnya, dipilih k centroid, yang merupakan titik pusat dari kluster, sebagai nilai awal. Setelah itu, jarak antara setiap data dan semua centroid yang ada dihitung. Kemudian, semua data dikelompokkan ke kluster yang paling mendekati. Proses iterasi akan berhenti ketika posisi centroid sudah diperbarui.

3.2. Flowchart Sistem Klasterisasi Tingkat Kriminalitas

Dalam gambar 2 sistem dimulai dengan proses masuk sebagai pengguna admin dan biasa. Setelah berhasil masuk, keduanya akan diarahkan ke halaman utama berdasarkan peran mereka. Pengguna admin memiliki hak penuh untuk mengatur data kriminalitas, penduduk, Indeks Pembangunan Manusia, kemiskinan, dan pengangguran menggunakan fitur CRUD, serta dapat melakukan klasterisasi menggunakan metode K-Means. Admin juga dapat logout untuk mengakhiri sesi. Sedangkan user hanya

dapat melihat hasil klasterisasi dan data yang telah diproses oleh admin dengan akses terbatas (read-only). Jika login gagal, pengguna akan kembali ke halaman login. User juga dapat logout kapan saja untuk keluar dari sistem.



Gambar 2. Flowchart Sistem Klasterisasi Tingkat Kriminalitas

3.3. Struktur Menu

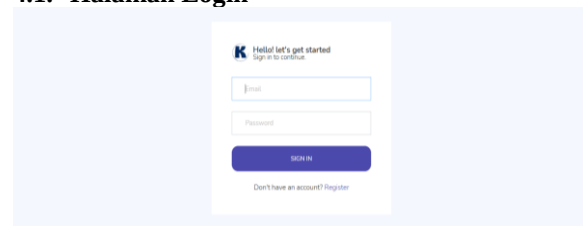


Gambar 3. Struktur Menu

Gambar 3 menunjukkan struktur menu pada website yang terdiri dari beberapa bagian, seperti dashboard, data kriminalitas, data penduduk, data IPM (Indeks Pembangunan Manusia), data kemiskinan, data pengangguran, dan klasterisasi. Pada menu data kriminalitas, data penduduk, data IPM, data kemiskinan, serta data pengangguran tersedia fitur CRUD yang digunakan untuk menambah, menampilkan, memperbarui, dan menghapus data.

4. HASIL DAN PEMBAHASAN

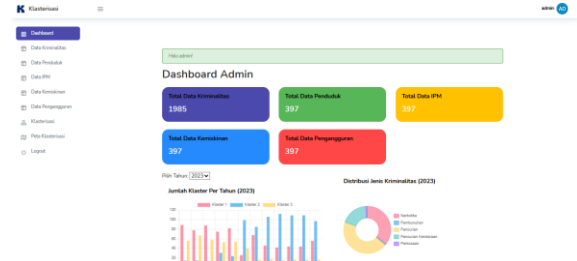
4.1. Halaman Login



Gambar 4. Halaman Login

Gambar 4 menunjukkan bagaimana halaman login muncul saat Anda pertama kali masuk ke website. Setelah itu, baik administrator maupun user harus memasukkan username atau email mereka untuk masuk dan mengakses menu, yang dapat ditemukan di website.

4.2. Dashboard



Gambar 5. Dashboard

Gambar 5 memperlihatkan halaman dashboard yang menampilkan jumlah informasi untuk berbagai jenis kejahatan, indeks pembangunan manusia, tingkat kemiskinan, angka pengangguran, dan data penduduk. Di samping itu, dasbor ini juga mempresentasikan grafik jumlah kluster setiap tahun dan jumlah informasi untuk masing-masing jenis kejahatan.

4.3. Data Kriminalitas

Tahun	Provinsi	Jenis Kriminalitas	Jumlah	Aksi
2023	Aceh	Perdagangan	15	[Edit] [Delete]
2023	Sulawesi Utara	Perdagangan	120	[Edit] [Delete]
2023	Sulawesi Barat	Perdagangan	15	[Edit] [Delete]
2023	Riau	Perdagangan	30	[Edit] [Delete]
2023	Jambi	Perdagangan	25	[Edit] [Delete]
2023	Sulawesi Selatan	Perdagangan	75	[Edit] [Delete]

Gambar 6. Menu Data Kriminalitas

Gambar 6 menampilkan halaman data kriminalitas dengan tabel berisi informasi dari setiap daerah. Fitur *Create*, *Read*, *Update*, dan *Delete* (CRUD) tersedia untuk pengelolaan data, ditambah dengan opsi impor file CSV atau Excel guna mempercepat input data dalam jumlah banyak.

4.4. Form Tambah Data Kriminalitas

Gambar 7. Form Tambah Data Kriminalitas

Gambar 7 menampilkan halaman form tambah data kriminalitas yang dilengkapi dengan input teks dan dropdown untuk mengisi detail data, sehingga memudahkan proses penambahan data ke tabel kriminalitas.

4.5. Data Penduduk

Tahun	Provinsi	Jumlah	Aksi
2023	Aceh	540215	[Edit] [Delete]
2023	Sulawesi Utara	125884	[Edit] [Delete]
2023	Sulawesi Barat	87072	[Edit] [Delete]
2023	Riau	464029	[Edit] [Delete]
2023	Jambi	387912	[Edit] [Delete]
2023	Sulawesi Selatan	874015	[Edit] [Delete]

Gambar 8. Menu Data Penduduk

Gambar 8 menunjukkan halaman data penduduk yang berisi tabel informasi tiap wilayah. Halaman ini menyediakan fitur *Create*, *Read*, *Update*, dan *Delete* (CRUD) serta impor file CSV atau Excel untuk memudahkan pengelolaan data skala besar.

4.6. Form Tambah Data Penduduk.

Gambar 9. Form Tambah Data Penduduk

Gambar 9 menunjukkan form tambah data penduduk dengan input teks dan dropdown untuk memudahkan penambahan data. Desain form ini membuat pengisian data lebih terstruktur dan efisien.

4.7. Data IPM (Indeks Pembangunan Manusia).

Tahun	Provinsi	Jumlah	Aksi
2023	Aceh	74.7	[Edit] [Delete]
2023	Sulawesi Utara	75.13	[Edit] [Delete]
2023	Sulawesi Barat	75.04	[Edit] [Delete]
2023	Riau	75.05	[Edit] [Delete]
2023	Jambi	75.73	[Edit] [Delete]
2023	Sulawesi Selatan	75.38	[Edit] [Delete]

Gambar 10 Menu Data IPM

Gambar 10 menunjukkan halaman data IPM (Indeks Pembangunan Manusia) dengan tabel informasi tiap wilayah. Tersedia fitur *Create*, *Read*, *Update*, dan *Delete* (CRUD) serta impor data CSV atau Excel untuk mempermudah pengelolaan data dalam jumlah besar.

4.8. Form Tambah Data IPM (Indeks Pembangunan Manusia)

Gambar 11. Form Tambah Data IPM.

Gambar 11 menampilkan halaman form tambah data IPM (Indeks Pembangunan Manusia) yang dilengkapi dengan input teks dan dropdown untuk menambahkan data ke tabel IPM.

4.9. Data Kemiskinan

Gambar 12. Menu Data Kemiskinan

Gambar 12 menampilkan halaman data kemiskinan dengan tabel data tiap wilayah. Fitur *Create*, *Read*, *Update*, dan *Delete* (CRUD) serta *import* data CSV atau *Excel* tersedia untuk memudahkan pengelolaan data dalam jumlah besar.

4.10. Form Tambah Data Kemiskinan

Gambar 13. Form Tambah Data Kemiskinan

Gambar 13 menunjukkan form tambah data kemiskinan dengan input teks dan *dropdown* untuk menambahkan data ke tabel. Desain form ini membuat proses penambahan data lebih mudah dan terstruktur.

4.11. Data Pengangguran

Tahun	Provinsi	Jumlah	Aksi
2023	Aceh	0.05	Tambah Hapus
2023	Sumatera Utara	0.08	Tambah Hapus
2023	Sumatera Barat	0.04	Tambah Hapus
2023	Riau	0.20	Tambah Hapus
2023	Jambi	0.03	Tambah Hapus
2023	Sumatera Selatan	0.11	Tambah Hapus

Gambar 14. Halaman Data Pengangguran

Gambar 14 menampilkan halaman data pengangguran yang berisi tabel informasi terkait tingkat pengangguran di berbagai wilayah. Halaman ini mendukung fitur *Create*, *Read*, *Update*, dan *Delete* (CRUD), yang memudahkan pengguna dalam mengelola data. Selain itu, tersedia pula opsi untuk mengimpor data melalui file berformat CSV atau Excel, sehingga memungkinkan penambahan data dalam jumlah besar dengan cepat dan efisien.

4.12. Form Tambah Data Pengangguran

Gambar 15. Form Tambah Data Pengangguran

Gambar 15 mengindikasikan form tambah data pengangguran dengan input teks dan *dropdown* untuk menambahkan data ke tabel. Form ini mempermudah proses pengisian data agar lebih teratur dan efisien.

4.13. Klasterisasi

Tahun	Provinsi	Jumlah Pengangguran	Jumlah Kriminalitas	IPM	Kemiskinan	Pengangguran	Tindakan
2023	Aceh	0.05	0.05	0.05	0.05	0.05	Tambah Hapus
2023	Sumatera Utara	0.08	0.08	0.08	0.08	0.08	Tambah Hapus
2023	Sumatera Barat	0.04	0.04	0.04	0.04	0.04	Tambah Hapus
2023	Riau	0.20	0.20	0.20	0.20	0.20	Tambah Hapus
2023	Jambi	0.03	0.03	0.03	0.03	0.03	Tambah Hapus
2023	Sumatera Selatan	0.11	0.11	0.11	0.11	0.11	Tambah Hapus
2023	Bengkulu	0.03	0.03	0.03	0.03	0.03	Tambah Hapus
2023	Lampung	0.06	0.06	0.06	0.06	0.06	Tambah Hapus
2023	Kabupaten Bengkulu Tengah	0.05	0.05	0.05	0.05	0.05	Tambah Hapus
2023	Kabupaten Pesisir	0.05	0.05	0.05	0.05	0.05	Tambah Hapus

Gambar 16. Menu Klasterisasi

Gambar 16 menampilkan halaman klasterisasi yang digunakan untuk menerapkan metode K-Means pada data kriminalitas, penduduk, IPM, kemiskinan, dan pengangguran. Hasil proses klasterisasi ditampilkan dalam bentuk tabel yang berisi pembagian data ke dalam masing-masing cluster. Halaman ini dirancang untuk menampilkan hasil pengelompokan data secara terstruktur dan jelas.

4.14. Peta Klasterisasi



Gambar 17. Halaman Peta Klasterisasi

Pada Gambar 17 merupakan halaman Peta Klasterisasi yang menampilkan hasil klasterisasi dalam bentuk visualisasi peta. Peta ini bertujuan untuk memberikan gambaran geografis terhadap persebaran klaster di masing-masing provinsi di Indonesia. Setiap provinsi diberi warna berdasarkan tingkat kriminalitasnya

4.15. Pengujian Metode

Pengujian metode K-Means dilakukan menggunakan *Silhouette Score* karena metrik ini dapat mengevaluasi kualitas klasterisasi tanpa memerlukan data berlabel. *Silhouette Score* menilai seberapa dekat suatu titik data dengan cluster-nya sendiri (kohesi), serta seberapa jauh dari cluster lain yang paling dekat (separasi), sehingga dapat menunjukkan apakah hasil klasterisasi sudah membentuk kelompok yang kompak dan saling terpisah dengan baik.

1. Menghitung a(i) : Range jarak antara titik data i dan seluruh titik data yang lain pada *cluster*-nya sendiri.

Cluster 1

Sumatera Barat = 0,12	Jambi = 0,20	...	Papua Barat = 0,73
-----------------------	--------------	-----	--------------------

$$(a) \text{ Aceh} = (0,12 + 0,42 + \dots + 0,73) / 10 = 0,37$$

Penghitungan a yang serupa dilaksanakan pada seluruh titik di cluster 1, berikut merupakan hasil a(Aceh) dan yang yang lain di *cluster* 1.

(a) Aceh = 0,37	(a) Banten = 0,48
(a) Sumatera Barat = 0,35	(a) Kalimantan Timur = 0,41
(a) Jambi = 0,37	(a) Sulawesi Utara = 0,42
(a) Kepulauan Riau = 0,47	(a) Maluku = 0,41
(a) DKI Jakarta = 0,70	(a) Papua = 0,89

Cluster 2

(a) Riau = 0,36	(a) Kalimantan Tengah = 0,32
(a) Sumatera Selatan = 0,61	(a) Kalimantan Selatan = 0,38
(a) Bengkulu = 0,31	(a) Kalimantan Utara = 0,36
(a) Lampung = 0,39	(a) Sulawesi Tengah = 0,33
(a) Kepulauan Bangka Belitung = 0,42	(a) Sulawesi Selatan = 0,54
(a) DI Yogyakarta = 0,52	(a) Sulawesi Tenggara = 0,31
(a) Bali = 0,42	(a) Gorontalo = 0,35
(a) Nusa Tenggara Barat = 0,33	(a) Sulawesi Barat = 0,43
(a) Nusa Tenggara Timur = 0,54	(a) Maluku Utara = 0,38
(a) Kalimantan Barat = 0,41	(a) Papua Barat = 0,42

Cluster 3

$$(a) \text{ Sumatera Utara} = 0,97$$

$$(a) \text{ Jawa Barat} = 0,74$$

$$(a) \text{ Jawa Tengah} = 0,67$$

$$(a) \text{ Jawa Timur} = 0,70$$

2. Menghitung b(i) : Range jarak antara titik data i dengan semua titik dari *cluster* terdekat.

Data klaster 1 ke setiap titik di klaster 2

$$\text{Aceh ke Riau} = 0,48$$

$$\text{Aceh ke Sumatera Selatan} = 0,62$$

$$\text{Aceh ke Papua Barat} = 0,59$$

$$b(\text{Aceh}) = (0,48 + 0,62 + \dots + 0,59) / 20 = 0,56$$

b(i) untuk Cluster 1

(b) Aceh = 0,56	(b) Banten = 0,82
(b) Sumatera Barat = 0,54	(b) Kalimantan Timur = 0,50
(b) Jambi = 0,46	(b) Sulawesi Utara = 0,59
(b) Kepulauan Riau = 0,75	(b) Maluku = 0,60
(b) DKI Jakarta = 0,90	(b) Papua = 0,97

b(i) untuk Cluster 2

(b) Riau = 0,51	(b) Kalimantan Tengah = 0,53
(b) Sumatera Selatan = 0,73	(b) Kalimantan Selatan = 0,52
(b) Bengkulu = 0,63	(b) Kalimantan Utara = 0,58
(b) Lampung = 0,55	(b) Sulawesi Tengah = 0,73
(b) Kepulauan Bangka Belitung = 0,51	(b) Sulawesi Selatan = 0,64
(b) DI Yogyakarta = 0,66	(b) Sulawesi Tenggara = 0,68
(b) Bali = 0,75	(b) Gorontalo = 0,74
(b) Nusa Tenggara Barat = 0,74	(b) Sulawesi Barat = 0,87
(b) Nusa Tenggara Timur = 0,93	(b) Maluku Utara = 0,57
(b) Kalimantan Barat = 0,88	(b) Papua Barat = 0,64

b(i) untuk Cluster 3

$$(b) \text{ Sumatera Utara} = 0,97$$

$$(b) \text{ Jawa Barat} = 0,74$$

$$(b) \text{ Jawa Tengah} = 0,67$$

$$(b) \text{ Jawa Timur} = 0,70$$

3. Menghitung s(i) : Skor *silhouette* untuk titik data i, perhitungan dimulai dari Aceh s(i) menggunakan persamaan (2).

$$s(\text{Aceh}) = \frac{0,56 - 0,37}{\max(0,37 - 0,56)} = 0,35$$

Selanjutnya, nilai a(i) dan b(i) dapat diperoleh untuk semua s(i) bagi seluruh titik data hasil clustering.

Tabel 1. Silhouette Score

Klaster	Provinsi	a(i)	b(i)	SC
Klaster 1	Aceh	0.37	0.56	0.35
	Sumatera Barat	0.35	0.54	0.36
	Jambi	0.37	0.46	0.19
	Kepulauan Riau	0.47	0.75	0.37
	DKI Jakarta	0.70	0.90	0.22
	Banten	0.48	0.82	0.42
	Kalimantan Timur	0.41	0.50	0.17
	Sulawesi Utara	0.42	0.59	0.28
	Maluku	0.41	0.60	0.32
Klaster 2	Papua	0.89	0.97	0.08
	Riau	0.36	0.51	0.29
	Sumatera Selatan	0.61	0.73	0.17
	Bengkulu	0.31	0.63	0.51
	Lampung	0.39	0.55	0.30
	Kepulauan Bangka Belitung	0.42	0.51	0.18
	DI Yogyakarta	0.52	0.66	0.21
	Bali	0.42	0.75	0.44
	Nusa Tenggara Barat	0.33	0.74	0.55
	Nusa Tenggara Timur	0.54	0.93	0.42
	Kalimantan Barat	0.41	0.88	0.53
	Kalimantan Tengah	0.32	0.53	0.40
	Kalimantan Selatan	0.38	0.52	0.27
	Kalimantan Utara	0.36	0.58	0.39
	Sulawesi Tengah	0.33	0.73	0.54
	Sulawesi Selatan	0.54	0.64	0.16
	Sulawesi Tenggara	0.31	0.68	0.55
	Gorontalo	0.35	0.74	0.52
	Sulawesi Barat	0.43	0.87	0.51
	Maluku Utara	0.38	0.57	0.33
Klaster 3	Papua Barat	0.42	0.64	0.34
	Sumatera Utara	0.97	0.86	-0.11
	Jawa Barat	0.74	1.35	0.45
	Jawa Tengah	0.67	1.14	0.41
	Jawa Timur	0.70	1.45	0.52

4. Rata – rata Silhouette Score s(i)

Cluster 1

$$\frac{0,35 + 0,36 + \dots + 0,08}{10} = 0,28$$

Pada klaster 1, skor Silhouette rata-rata adalah 0,28. Nilai ini menunjukkan bahwa beberapa provinsi dalam klaster 1 memiliki kualitas pengelompokan yang buruk, yang berarti bahwa data dalam klaster ini mungkin kurang sesuai.

Cluster 2

$$\frac{0,29 + 0,17 + \dots + 0,34}{20} = 0,38$$

Klaster 2 memiliki skor Silhouette rata-rata 0,38. Hasil ini menunjukkan bahwa, dibandingkan dengan klaster 1 dan 3, beberapa provinsi dalam klaster ini memiliki kualitas klusterisasi yang lebih baik.

Cluster 3

$$\frac{-0,11 + 0,45 + \dots + 0,52}{4} = 0,32$$

Pada klaster 3, skor Silhouette rata-rata adalah 0,32. Nilai-nilai ini menunjukkan bahwa beberapa provinsi dalam klaster ini memiliki

kualitas klusterisasi yang cukup baik, tetapi beberapa provinsi menerima nilai yang buruk, yang menunjukkan bahwa data ini lebih dekat dengan klaster lain daripada klaster 3.

Silhouette Score Keseluruhan

$$\frac{0,35 + 0,36 + \dots + 0,52}{34} = 0,34$$

Hasil klusterisasi secara keseluruhan cukup baik, dengan rata-rata *Silhouette Score* 0,34, menunjukkan bahwa sebagian besar data telah dikelompokkan ke dalam klaster yang tepat, tetapi beberapa data belum ditempatkan dengan baik karena nilai *Silhouette Score* negatif.

5. KESIMPULAN DAN SARAN

Sistem klusterisasi tingkat kriminalitas di Indonesia dengan algoritma K-Means berhasil mengelompokkan provinsi berdasarkan kemiripan data sosial dan kriminalitas. Hasil pengujian rata-rata *Silhouette Score* sebesar 0,34 menunjukkan kualitas klusterisasi yang cukup baik, dengan Cluster 3 sebagai yang terbaik 0,38.

Untuk pengembangan selanjutnya, disarankan agar data diperluas hingga mencakup seluruh kota di setiap provinsi guna memperluas cakupan analisis. Selain itu, penggunaan metode K-Means++ direkomendasikan karena pemilihan centroid awal yang lebih akurat. Evaluasi di masa mendatang juga sebaiknya mempertimbangkan lebih banyak indikator sosial untuk hasil yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] M. A. Z. Ramadhan, R. A. Saputra dan J. Nangi, "Klusterisasi Daerah Rawan Kriminalitas di Sulawesi Tenggara Menggunakan Metode K-Means Clustering," *JOINTER: Journal Of Informatics Engineering*, vol. 5, no. 1, pp. 1-6, 2024
- [2] H. Syaputra, D. P. Prahastiwi, R. Dewi dan N. Sari, *Statistik Kriminal 2024*, BPS Indonesia, 2024
- [3] "Statistik Kriminal 2019," *BPS(Badan Pusat Statistik)*, 2019
- [4] D. Yuliyanti dan M. Martanto, "Clustering Tingkat Kejahatan Kriminal Menggunakan Metode K-Means di Wilayah Kabupaten Cirebon," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3509-3514, 2023
- [5] D. Sutejo, Y. A. Pranoto dan H. Z. Zahro, "Sistem Informasi Geografis Pengelompokan Tingkat Kriminalitas Kota Malang Menggunakan Metode K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 4, no. 1, pp. 356-363, 2020
- [6] I. D. Anjani dan A. Bahtiar, "Penerapan Algoritma K-Means Clustering untuk mengelompokkan penerima bantuan sosial Tunai (BST) di Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2743-2747, 2024

- [7] D. Gultom, H. D. Tampubolon, F. I. R. H. Zer dan D. Hartama, "Penerapan Algoritma K-Means untuk mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *JurTi (Jurnal Teknologi Informasi)*, vol. 4, no. 1, pp. 146-151, 2020
- [8] D. A. Azhari, Y. Maulita dan S. Ramadani, "Pengelompokan Data Kriminal untuk Menentukan Pola Rawan Tindak Kriminal Menggunakan Algoritma K-Means(Studi Kasus: Polsek Hamparan Perak)," *Polygon : Jurnal Ilmu Komputer dan Ilmu Pengetahuan Alam*, vol. 2, no. 5, pp. 122-138, 2024
- [9] A. G. Fadilah, F. S. Wahyuni dan S. Achmadi, "IMPLEMENTASI METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN KELAS BERDASARKAN PEMAHAMAN SISWA PADA BIMBINGAN BELAJAR MATEMATIKA SASCHIO BANYUWANGI," *JATI(Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 1072-1077, 2022
- [10] E. Muzilin dan M. Martanto, "ANALISIS K-MEANS CLUSTERING UNTUK MENGIDENTIFIKASI KEBERADAAN BADAN USAHA MILIK DESA DI JAWA BARAT," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1882-1888, 2024
- [11] B. W. Nugraha, A. Mahmudi dan F. S. Wahyuni, "PENERAPAN METODE K-MEANS UNTUK PENGELOMPOKAN TINGKAT KEPUASAN PENGGUNA LULUSAN PADA TRACER STUDY PUSAT KARIR ITN MALANG," *JATI(Jurnal Mahasiswa Teknik Informatika)*, vol. 5, no. No. 2, pp. 684-692, 2021
- [12] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan," *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, vol. Vol. 06, no. No. 01, pp. 60-74, 2024
- [13] S. M. Haq, D. Rudhistiar dan R. P. Prasetya, "PENERAPAN DATA MINING MENGGUNAKAN K-MEANS CLUSTERING UNTUK PENGELOMPOKAN STATUS KEMISKINAN DI PAPUA," *JATI (Journal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 3551-3559, 2025
- [14] M. R. Dirgantara, S. Syahputri, A. Hasibuan dan N. N, "Pengenalan Database Management System (DBMS)," *Madani: Jurnal Ilmiah Multidisiplin*, vol. 1, no. 6, pp. 300-306, 2023
- [15] N. A. Alifia, F. S. Setiawan dan D. Rudhistiar, "PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM PERINGATAN DINI RESIKO KEGAGALAN SISWA PADA MATA PELAJARAN BAHASA INDONESIA (STUDI KASUS: SMP NEGERI 2 MALANG)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. Vol. 8 , no. No. 2, pp. 1175-1181, 2024
- [16] W. P. Priyadi, J. D. Irawan dan A. Faisol, "Penerapan Data Mining Untuk Clustering Wilayah Produksi Padi Menggunakan Metode K-Means (Studi Kasus : Wilayah Jawa Timur)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 8381-8388, 2024