



Analisis *Cluster* dengan Metode SOM untuk Pengelompokan Provinsi di Indonesia berdasarkan Indikator Kriminalitas

Cluster Analysis using SOM Method for Grouping of Provinces in Indonesia based on National Crime Indicators

Mardhani Dwi Novianto, Prodi Matematika FMIPA UNY
Sri Andayani*, Prodi Matematika FMIPA UNY
*e-mail: andayani@uny.ac.id

Abstrak

Kriminalitas atau kejahatan adalah suatu perbuatan yang dapat mengakibatkan masalah bagi masyarakat. Salah satu teknik analisis untuk mengatasi masalah ini adalah analisis *cluster*. Hasil analisis ini dapat dijadikan evaluasi bagi pemerintah dalam memperhatikan kriminalitas yang terjadi. Penelitian ini bertujuan untuk mendeskripsikan hasil analisis *cluster* dan karakteristik dari hasil analisis *cluster* dengan metode *Self-Organizing Maps* (SOM) untuk pengelompokan provinsi di Indonesia berdasarkan indikator kriminalitas nasional. Penelitian ini menggunakan data indikator kriminalitas nasional tahun 2021 yang diperoleh dari laman resmi Badan Pusat Statistik Indonesia. Data yang digunakan terdiri dari 34 provinsi dan 7 variabel. Hasil penelitian menunjukkan bahwa 5 *cluster* adalah jumlah *cluster* optimal dengan nilai validasi DBI 0,3793. Setiap *cluster* memiliki karakteristiknya tersendiri. *Cluster* ke-1 memiliki rata-rata jumlah kejahatan ringan, sedang, berat, dan jumlah kejahatan yang diselesaikan paling rendah, serta selang waktu terjadinya kejahatan paling lama. *Cluster* ke-2 memiliki rata-rata jumlah kejahatan sedang dan berat paling tinggi, serta selang waktu terjadinya kejahatan paling cepat. *Cluster* ke-3 memiliki rata-rata jumlah kejahatan ringan paling tinggi. *Cluster* ke-4 memiliki rata-rata persentase penyelesaian kejahatan dan risiko penduduk terkena kejahatan paling rendah. *Cluster* ke-5 memiliki rata-rata jumlah kejahatan yang diselesaikan, persentase penyelesaian kejahatan, dan risiko penduduk terkena kejahatan paling tinggi.

Kata kunci: Indikator Kriminalitas Nasional, analisis *cluster*, *Self-Organizing Maps*, *Davies Bouldin Index*

Abstract

Crime is an act that can cause problems for society. One of the analytical techniques to overcome this problem is cluster analysis. The results of this analysis can be used as an evaluation for the government in paying attention to crime in Indonesia. This study aims to describe the results of cluster analysis and the characteristics of the results of cluster analysis using Self-Organizing Maps (SOM) method for grouping of provinces in Indonesia based on national crime indicators. This study uses national crime indicators data for 2021 obtained from the official website of the Indonesia Central Statistics Agency (BPS). The data used consist of 34 provinces and 7 variables. The results showed that 5 clusters is the optimal number of clusters with a DBI validation value of 0.3793. Each cluster has its own characteristics. First cluster has the lowest average number of minor, moderate, serious crimes, and the number of crime cleared, but the longest time interval between crimes (crime clock). Second cluster has the highest average number of moderate and serious crime, but the shortest time interval between crimes (crime clock). Third cluster has the highest average number of minor crime. Fourth cluster has the lowest average of clearance rate and crime rate. Fifth cluster has the highest average number of crime cleared, clearance rate, and crime rate.

Keywords: *National Crime Indicators, cluster analysis, Self-Organizing Maps, Davies Bouldin Index*

PENDAHULUAN

Kriminalitas atau kejahatan didefinisikan sebagai suatu perbuatan yang dapat berakibat pada timbulnya masalah-masalah dan keresahan bagi kehidupan masyarakat (Lestari, 2018). Tindak kejahatan/kriminalitas atau pelanggaran merupakan perbuatan seseorang yang dapat diancam hukuman berdasarkan Kitab Undang-Undang Hukum Pidana (KUHP) atau Undang-Undang serta peraturan lainnya yang berlaku di Indonesia (Statistik, 2021). Di negara hukum, konsep kriminalitas dipandang sebagai suatu perbuatan yang dilarang oleh undang-undang dan diancam dengan sanksi pidana bagi yang melanggar larangan tersebut (Dulkiah, 2020).

Berdasarkan data yang diperoleh dari *Numbeo* (<https://www.numbeo.com/>), tingkat kriminalitas di Indonesia pada pertengahan tahun 2022 menduduki peringkat keempat di antara negara-negara Asia Tenggara dengan indeks kejahatan sebesar 46.1 dari skala 100. Data yang dipublikasikan oleh Badan Pusat Statistika dalam Statistik Kriminal tahun 2022 (<https://www.bps.go.id/publication.html>) memperlihatkan bahwa jumlah kejahatan (*crime total*) pada tahun 2019 sebanyak 269,324 kasus, menurun menjadi sebanyak 247,218 kasus pada tahun 2020 dan menurun pada tahun 2021 menjadi 239,481 kasus. Pada tahun 2021, jumlah kejahatan yang diselesaikan sebanyak 172,065 kasus, berarti ada 71,85% kasus kejahatan yang diselesaikan. Selama periode tahun 2019-2021, jumlah kejadian kejahatan atau tindak kriminalitas di Indonesia cenderung menurun, tetapi belum semua kasus dapat diselesaikan. Dengan melihat kriminalitas di Indonesia, hal tersebut dapat menjadi suatu perhatian karena dapat meresahkan dan mempengaruhi rasa aman bagi kehidupan masyarakat, terlebih lagi, kejahatan atau kriminalitas dapat terjadi di mana pun dan kapan pun. Kejahatan masih menjadi masalah yang sulit untuk ditangani secara tuntas.

Dalam statistik kriminal terdapat beberapa indikator yang biasa digunakan untuk mengukur kejahatan dari sisi makro. Dalam konteks makro, terdapat indikator angka jumlah kejahatan, angka kejahatan per 100.000 penduduk (*crime rate*), dan selang waktu terjadinya suatu tindak kejahatan (*crime clock*). Semakin tinggi angka jumlah kejahatan dan angka kejahatan per 100.000 penduduk, serta semakin cepat selang waktu terjadinya suatu tindak kejahatan, maka tingkat kerawanan akan kejahatan di suatu daerah semakin tinggi pula, dan sebaliknya (Statistik, 2022). Selain itu, juga terdapat indikator yang lain, yaitu jumlah kejahatan yang diselesaikan (*crime cleared*) dan persentase penyelesaian kejahatan (*clearance rate*). Semakin tinggi kedua indikator tersebut, maka semakin rendah tingkat kerawanan akan kejahatan di suatu daerah dan sebaliknya. Indikator kriminalitas dapat menjadi gambaran dalam menganalisis kondisi kriminalitas dari setiap provinsi di Indonesia.

Adanya kejadian kriminalitas atau kejahatan di Indonesia menunjukkan perlunya suatu upaya untuk menangani masalah ini. Upaya yang dapat dilakukan adalah melakukan suatu pengelompokan terhadap provinsi-provinsi di Indonesia yang ditinjau dari indikator kriminalitas. Hal ini tidak hanya dapat membantu masyarakat untuk tetap waspada terhadap kriminalitas atau kejahatan, tetapi juga dapat menjadi evaluasi bagi pemerintah dan berpengaruh terhadap kebijakan yang nantinya akan diterapkan oleh pemerintah. Salah satu teknik analisis yang dapat digunakan untuk menangani masalah kriminalitas atau kejahatan adalah analisis *cluster*. Analisis *cluster* merupakan suatu teknik analisis yang berguna untuk meringkas data yang dapat dilakukan dengan jalan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu di antara objek-objek yang hendak diteliti (Sunarso, 2008).

Pada penelitian ini, dilakukan analisis *cluster* dengan menggunakan metode *Self-Organizing Maps* (SOM). Menurut Yang, Ouyang, & Shi (2012), *Self-Organizing Maps* termasuk dalam analisis *cluster* yang populer dari teknik *neural network*. Teknik tersebut banyak diterapkan untuk memecahkan berbagai masalah, termasuk dalam analisis *cluster* karena memiliki keunggulan berupa kemampuan belajar yang baik, kemampuan adaptasi, kemudahan implementasi, kecepatan, dan fleksibilitas. Menurut Wahyudi, Suhartono, & Pramunendar (2015), *Self-Organizing Maps* merupakan salah satu algoritma dalam *Artificial*

Neural Network yang termasuk ke dalam jenis pola pembelajaran *unsupervised*. SOM juga memiliki keunggulan tahan terhadap data noise yang disebut *outlier*, serta dapat mengatasi data yang besar (Bustamam et al., 2018).

Khotija (2022) melakukan penelitian terkait pengelompokan tindak kriminal di Indonesia menggunakan metode *Self-Organizing Maps* yang dibandingkan dengan *K-Means*. Pada penelitian tersebut menggunakan data dengan jumlah 22 variabel yang merupakan jenis tindak kriminal, yaitu pembunuhan, penganiayaan berat, penganiayaan ringan, KDRT, pemerkosaan, pencabulan, penculikan, mempekerjakan anak di bawah umur, pencurian dengan kekerasan, pencurian dengan kekerasan (senpi), pencurian dengan kekerasan (sajam), pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, penadahan, pengrusakan/penghancuran barang, pembakaran dengan sengaja, narkoba dan psikotropika, penipuan, penggelapan, korupsi, dan terhadap ketertiban umum. Diperoleh hasil sebanyak 3 *cluster* untuk masing-masing metode dengan anggota *cluster* yang berbeda-beda. Berdasarkan nilai *Davies Bouldin Index*, diperoleh hasil pengelompokan terbaik menggunakan metode SOM yaitu sebesar 0,08 lebih kecil dibandingkan nilai *K-Means* yaitu sebesar 0,40. Penelitian tersebut sebagai penelitian terdahulu dan referensi untuk penelitian ini dalam penggunaan metode *Self-Organizing Maps*, tetapi dengan penggunaan variabel yang berbeda.

SOM dianggap sebagai bentuk *clustering k-means* yang dibatasi secara spasial (Wehrens, 2007). Menurut Choudhary (2019), dalam beberapa kasus, jika inisialisasi pusat *cluster* tidak sesuai, *K-Means* dapat menghasilkan *cluster* yang buruk. Algoritma *K-Means++* dapat membantu kondisi tersebut dengan menentukan prosedur untuk menginisialisasi pusat *cluster*. Algoritma ini mengoptimalkan langkah dimana secara acak memilih pusat *cluster*, dan memungkinkan menemukan solusi yang kompetitif dengan solusi *K-Means* yang optimal saat menggunakan inisialisasi *K-Means++*. Berdasarkan hal tersebut, penentuan nilai bobot neuron awal pada SOM juga dapat menerapkan algoritma *K-Means++*. Menurut Tempola, Muhammad, & Mubarak (2020), validasi *cluster* digunakan untuk memutuskan *cluster* terbaik pada dataset yang sedang diuji. Untuk validasi *cluster* ada beberapa cara, salah satunya dengan *Davies-Bouldin Index*. *Davies-Bouldin Index (DBI)* adalah ukuran untuk mengevaluasi kinerja *cluster* (Davies & Bouldin, 1979). Jumlah *cluster* dapat ditentukan sebanyak dua, tiga, empat, dan seterusnya dimana setiap *cluster* memiliki karakteristik yang sama. Jumlah *cluster* yang optimal ditunjukkan dengan nilai DBI yang paling rendah (P. et al., 2014).

Berdasarkan uraian di atas, pada penelitian ini akan dilakukan analisis *cluster* untuk pengelompokan terhadap provinsi-provinsi di Indonesia yang ditinjau dari indikator kriminalitas karena indikator ini dapat digunakan untuk menganalisis kondisi kriminalitas di Indonesia. Metode yang digunakan adalah *Self Organizing Maps* yang dinilai akan memberikan hasil yang terbaik dan akan diterapkan dengan mempertimbangkan keunggulan-keunggulannya antara lain termasuk salah satu algoritma dalam *Artificial Neural Network* yang mana memiliki keunggulan berupa kemampuan belajar yang baik, kemampuan adaptasi, kemudahan implementasi, kecepatan, dan fleksibilitas. Selain itu, juga tahan terhadap data noise yang disebut *outlier*, serta dapat mengatasi data yang besar. Data yang akan digunakan adalah data indikator kriminalitas nasional tahun 2021. Sementara untuk memvalidasi jumlah *cluster* yang optimal digunakan *Davies-Bouldin Index*.

METODE

Penelitian ini menggunakan data indikator kriminalitas nasional tahun 2021 yang diperoleh dari laman resmi Badan Pusat Statistik Indonesia dalam publikasinya yang berjudul Statistik Kriminal 2022 (<https://www.bps.go.id/publication.html>). Data yang digunakan terdiri dari 34 provinsi dan 7 variabel berupa indikator kriminalitas nasional yang dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Indikator	Definisi
Jumlah Kejahatan Tingkat Ringan	Jumlah perkara kejahatan yang diterima kepolisian dari laporan masyarakat dengan ancaman hukuman penjara atau kurungan paling lama tiga bulan
Jumlah Kejahatan Tingkat Sedang	Jumlah perkara kejahatan yang diterima kepolisian dari laporan masyarakat dengan ancaman hukuman lebih dari 3 bulan sampai dengan 5 tahun penjara
Jumlah Kejahatan Tingkat Berat	Jumlah perkara kejahatan yang diterima kepolisian dari laporan masyarakat dengan ancaman hukuman paling sedikit 5 tahun penjara atau seumur hidup atau hukuman mati
Jumlah Kejahatan yang Diselesaikan	Jumlah seluruh kejahatan yang diselesaikan oleh kepolisian pada satu tahun
Persentase Penyelesaian Kejahatan	Persentase jumlah peristiwa kejahatan yang diselesaikan terhadap jumlah kejahatan yang dilaporkan
Selang Waktu Terjadinya Kejahatan	Selang waktu atau interval waktu terjadinya satu tindak kejahatan dengan kejahatan yang lain
Risiko Penduduk Terkena Kejahatan	Jumlah kejahatan setahun dibagi dengan jumlah penduduk tahun yang bersangkutan dikalikan 100.000

Adapun langkah-langkah dalam penelitian ini yaitu, pengumpulan data, analisis statistika deskriptif, data *preprocessing*, analisis *cluster* dengan *Self-Organizing Maps*, penentuan jumlah *cluster* optimal dengan *Davies Bouldin Index*, analisis karakteristik *cluster*, pemetaan *cluster*, dan simpulan. Tahap pertama diawali dengan pengumpulan data, kemudian dilakukan analisis statistika deskriptif. Pada hasil analisis ini, ditampilkan nilai rata-rata, standar deviasi, minimal, dan maksimal. Tahap selanjutnya adalah data *preprocessing* dengan melakukan pengecekan apakah terdapat nilai yang hilang atau pun terdapat pencilan (*outlier*) dari setiap variabel, kemudian dilakukan normalisasi data menggunakan *z-score*. Adapun *z-score* dapat dihitung dengan persamaan (1) berikut.

$$Z - score = \frac{x_{ij} - \bar{x}_i}{s_i} \quad (1)$$

dimana x_{ij} adalah sampel data, \bar{x}_i adalah rata-rata nilai dan s_i adalah standar deviasi.

Tahap selanjutnya adalah analisis *cluster* dengan metode *Self-Organizing Maps* (SOM). Adapun langkah-langkah atau tahap dari algoritma *Self-Organizing Maps* adalah sebagai berikut (Halim & Widodo, 2017).

1. Inisialisasi berupa bobot (w_{ij}) yang diperoleh secara acak untuk tiap node. Setelah bobot (w_{ij}) diberikan, maka jaringan diberikan *input* (x_i).
2. Setelah input diterima, jaringan akan melakukan perhitungan jarak *vector* $d_{(j)}$ yang didapat dengan menjumlahkan selisih antara vektor bobot (w_{ij}) dengan vektor input (x_i).

$$d_j = \sum_i (w_{ij} - x_i)^2 \quad (2)$$

3. Setelah jarak antara node diketahui, maka ditentukan nilai minimum dari perhitungan jarak vektor d_j . Tahap selanjutnya melakukan perubahan bobot.

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha[x_i - w_{ij}(\text{old})] \quad (3)$$

4. Pada proses untuk mendapatkan bobot baru memerlukan nilai *learning rate* (α) yaitu $0 \leq \alpha \leq 1$. Nilai *learning rate* pada setiap iterasi / *epoch* akan berkurang menjadi $\alpha(i + 1) = 0.5\alpha$.
5. Kondisi penghentian pengujian dilakukan dengan menghitung selisih antara bobot $w_{ij}(\text{new})$ dan $w_{ij}(\text{old})$, apabila nilai w_{ij} hanya berubah sedikit, berarti pengujian sudah mencapai konvergensi sehingga dapat dihentikan.

Sebagai tambahan pada langkah 1 tahap dari algoritma *Self-Organizing Maps*, inisialisasi berupa bobot (w_{ij}) dilakukan dengan algoritma *K-Means++*. Adapun langkah-langkah atau tahap dari algoritma *K-Means++* adalah sebagai berikut (Gupta, 2021).

1. Memilih inisialisasi pusat *cluster* pertama (c_1) secara acak dari data.
2. Memilih pusat *cluster* selanjutnya c_i , dengan memilih $c_i = x' \in X$ berdasarkan probabilitasnya dengan persamaan sebagai berikut.

$$p(x) = \frac{D(x')^2}{\sum_{x \in X} D(x)^2} \quad (4)$$

dimana $D(x) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$

Keterangan :

n = banyak data

x_i = titik data ke- i

c_i = pusat *cluster* ke- i

x' = kandidat pusat *cluster* yang dipilih dalam proses inisialisasi

X = dataset yang direpresentasikan sebagai matriks berukuran $n \times m$

$D(x')^2$ = jarak kuadrat antara kandidat pusat *cluster* x' dan pusat *cluster* yang telah dipilih

$D(x)^2$ = jarak kuadrat antara titik data dalam dataset X dan pusat *cluster* yang telah dipilih

$p(x)$ = probabilitas data

Memilih pusat *cluster* berikutnya (c_i) dari data dimana probabilitas memilih titik sebagai pusat *cluster* berbanding lurus dengan jarak dari *centroid* terdekat yang dipilih, artinya bahwa data yang memiliki probabilitas tertinggi atau dengan kata lain memiliki jarak maksimum dari *centroid* terdekat, maka akan dipilih sebagai pusat *cluster* berikutnya.

3. Lakukan langkah 2 dan 3 sampai diperoleh sejumlah k pusat *cluster*

Setelah diperoleh hasil dari analisis *cluster* dengan metode *Self-Organizing Maps*, dilakukan penentuan jumlah *cluster* optimal dengan *Davies Bouldin Index*. *Davies Bouldin Index* adalah sebuah ukuran untuk mengevaluasi kinerja *clustering* / pengelompokan (Davies & Bouldin, 1979). Semakin kecil nilai yang didapat, maka kemiripan data dalam satu kelompok akan semakin tinggi, dengan kata lain, jumlah *cluster* optimal ditunjukkan oleh nilai yang terkecil (Pamulang et al., 2021). Adapun langkah-langkah atau tahap dari *Davies Bouldin Index* adalah sebagai berikut (Septiani et al., 2022).

1. Menghitung *Sum of Square within Cluster* (SSW), untuk mengetahui kohesi/homogenitas. Kohesi merupakan keterikatan anggota *cluster* dalam satu *cluster*. Adapun persamaannya sebagai berikut.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (5)$$

Keterangan :

m_i = jumlah data yang berada dalam *cluster* ke-i

x = data dalam *cluster*

$d(x, c)$ = jarak data ke *centroid*

x_j = data pada *cluster* tersebut

c_i = *centroid cluster* ke-i

2. Menghitung *Sum of Square between Cluster* (SSB), untuk mengetahui separasi/heterogenitas. Separasi merupakan perbedaan antara satu *cluster* dengan *cluster* lainnya. Adapun persamaannya sebagai berikut.

$$SSB_{i,j} = d(c_i, c_j) \quad (6)$$

Keterangan :

c_i = *cluster* ke-i

c_j = *cluster* ke-j

$d(c_i, c_j)$ = jarak antara *centroid* satu dengan lainnya

3. Menghitung Rasio (R_{ij}), untuk mengetahui seberapa bagus nilai perbandingan *cluster* satu dengan *cluster* yang lainnya. Adapun persamaannya sebagai berikut.

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (7)$$

Keterangan :

R_{ij} = Rasio antar *cluster*

SSW_i = kohesi dari *cluster* ke-i

SSW_j = kohesi dari *cluster* ke-j

$SSB_{i,j}$ = separasi dari *cluster* ke-i dan ke-j

4. Menghitung *Davies Bouldin Index* (DBI). Adapun persamaannya sebagai berikut.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (8)$$

Keterangan :

k = jumlah *cluster* yang digunakan

Setelah diperoleh jumlah *cluster* yang optimal, dilakukan analisis karakteristik *cluster*. Tahap ini dilakukan dengan menghitung rata-rata variabel dari setiap *cluster*. Kemudian, dilakukan visualisasi dalam bentuk peta Indonesia untuk mempermudah pemahaman dan identifikasi dari *cluster* yang dibentuk. Tahap terakhir adalah simpulan dimana pada bagian ini menjawab permasalahan yang telah dirumuskan dalam penelitian dan memberikan saran yang berkaitan dengan meminimalisir permasalahan untuk ke depannya.

HASIL DAN PEMBAHASAN

Hasil

1. Analisis Statistika Deskriptif

Analisis statistika deskriptif dilakukan untuk mendapatkan gambaran distribusi data. Pada hasil analisis ini, ditampilkan nilai rata-rata, standar deviasi, minimal, dan maksimal. Hasil statistika deskriptif yang menggambarkan indikator kriminalitas nasional di Indonesia pada tahun 2021 ditampilkan dalam Tabel 2.

Tabel 2. Hasil Statistika Deskriptif

Variabel	Rata-rata	Standar deviasi	Minimal	Maksimal
X1	433,62	457,64	0	2192
X2	2633,47	2824,22	173	13974
X3	2384,15	2909,10	269	14527
X4	5060,74	6061,90	589	28518
X5	67,62	17,16	32,83	100
X6	9024,85	7465,30	863	32477
X7	130	66,98	15	289

2. Data Preprocessing

Data yang telah dikumpulkan, dilakukan pemrosesan terlebih dahulu sebelum dilakukan pembentukan *cluster*. Adapun tahap yang dilakukan adalah sebagai berikut.

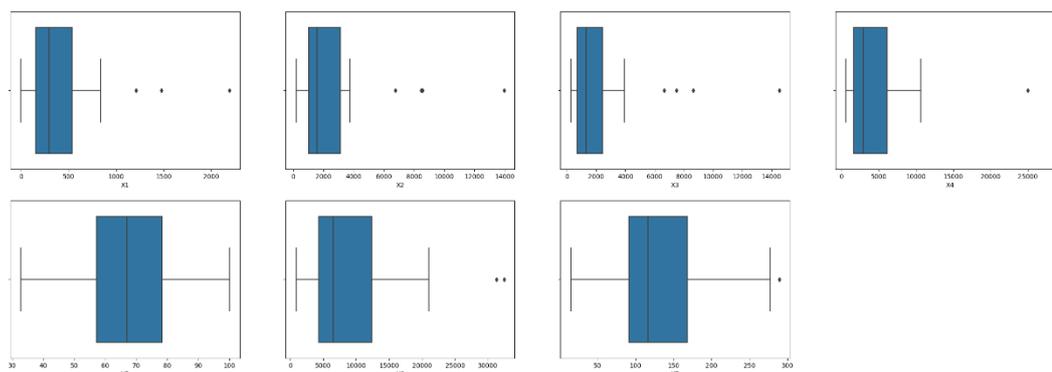
a. Data Cleaning

Pada tahap ini, dilakukan pengecekan apakah terdapat nilai yang hilang atau pun terdapat pencilan (*outlier*) dari setiap variabel yang disajikan pada Tabel 3 dan Gambar 1.

Tabel 3. Jumlah nilai yang hilang

	X1	X2	X3	X4	X5	X6	X7
Jumlah nilai yang hilang	0	0	0	0	0	0	0

Pada Tabel 3 menampilkan jumlah nilai yang hilang dari setiap variabel. Terdapat angka 0 pada setiap variabel, hal ini menunjukkan bahwa tidak terdapat nilai yang hilang. Selanjutnya, dilakukan pengecekan terhadap pencilan (*outlier*) dengan menggunakan *boxplot*.



Gambar 1. Pengecekan pencilan (*outlier*) dengan *boxplot*

Berdasarkan Gambar 1 dapat ditunjukkan bahwa terdapat pencilan pada data. Untuk mengatasi hal ini, data pencilan dapat dihapuskan atau dihilangkan. Akan tetapi dengan penghapusan atau penghilangan ini, dapat berpengaruh terhadap hasil atau

kesimpulan dari data yang miliki. Sehingga untuk mengatasi pencilan ini, dapat dilakukan transformasi data dengan normalisasi.

b. Normalisasi data

Pada tahap ini, dilakukan normalisasi data untuk mengatasi adanya pencilan pada data dan juga untuk mengubah data ke dalam skala yang teratur sehingga data lebih akurat. Pada penelitian ini, normalisasi data menjadi nilai *z-score*. Perhitungan dilakukan dengan persamaan (1), sebagai contoh akan ditunjukkan salah satu variabel sebelum dan sesudah normalisasi pada Tabel 4.

Tabel 4. Hasil Normalisasi Data

Provinsi	X1	
	Data Asli	Z-score
Aceh	675	0,5274
Bali	203	-0,5039
Kep Bangka Belitung	3	-0,9409
Banten	237	-0,4296
Bengkulu	58	-0,8208

3. Analisis *Cluster* dengan *Self-Organizing Maps*

Pada tahap ini, dilakukan analisis *cluster* atau *clustering* dengan metode *Self-Organizing Maps* untuk pengelompokan provinsi berdasarkan indikator kriminalitas nasional pada tahun 2021. Adapun parameter-parameter yang digunakan adalah sebagai berikut.

- a. Jumlah *cluster* (k) = 2, 3, 4, 5, 6, 7, 8, 9, dan 10
- b. *Learning rate* (α) = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; dan 0,9
- c. Jumlah iterasi = 1, 5, 10, 15, dan 20
- d. Untuk inialisasi bobot awal dilakukan dengan inialisasi *K-Means++*

Pada penelitian ini jumlah iterasi yang digunakan ditentukan berdasarkan hasil uji coba dengan beberapa parameter yang telah disebutkan. Hasil dari uji coba yang dilakukan disajikan pada Tabel 5.

Tabel 5. Hasil uji coba untuk mencari jumlah iterasi yang stabil ditinjau dari nilai DBI dengan jumlah *cluster* sebanyak 2 s.d. 10 *cluster* dan *learning rate* 0,1 s.d. 0,9

α	Jumlah Iterasi	Jumlah Cluster								
		2	3	4	5	6	7	8	9	10
0,1	1	0,5836	0,5094	0,4935	0,3793	0,7449	0,5973	0,6511	0,5900	0,6190
	5	0,5836	0,5094	0,4935	0,3793	0,7409	0,5629	0,6167	0,5736	0,5867
	10	0,5836	0,5094	0,4935	0,3793	0,7409	0,5629	0,6167	0,5736	0,5867
	15	0,5836	0,5094	0,4935	0,3793	0,7409	0,5629	0,6167	0,5736	0,5867
	20	0,5836	0,5094	0,4935	0,3793	0,7409	0,5629	0,6167	0,5736	0,5867
0,2	1	0,5836	0,5094	0,7226	0,5746	0,5767	0,5789	0,6281	0,5797	0,5480
	5	0,5836	0,5094	0,7226	0,5746	0,5802	0,5667	0,6167	0,5736	0,5371
	10	0,5836	0,5094	0,7226	0,5746	0,5802	0,5667	0,6167	0,5736	0,5371
	15	0,5836	0,5094	0,7226	0,5746	0,5802	0,5667	0,6167	0,5736	0,5371
	20	0,5836	0,5094	0,7226	0,5746	0,5802	0,5667	0,6167	0,5736	0,5371
0,3	1	0,5836	0,5094	0,4935	0,3793	0,5759	0,5688	0,6220	0,5777	0,5885
	5	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6187	0,5736	0,5867
	10	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6187	0,5736	0,5867
	15	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6187	0,5736	0,5867
	20	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6187	0,5736	0,5867
0,4	1	0,5836	0,5094	1,0069	0,7892	0,5759	0,5688	0,6220	0,5777	0,5885
	5	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,5867

	10	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,5867
	15	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,5867
	20	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,5867
0,5	1	0,5836	0,5094	1,0069	0,7892	0,5759	0,5688	0,6220	0,5777	0,6310
	5	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,6242
	10	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,6242
	15	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,6242
	20	0,5836	0,8275	0,8223	0,6483	0,5802	0,5667	0,7296	0,6859	0,6242
0,6	1	0,5836	0,5094	0,4935	0,3793	0,6686	0,5688	0,7030	0,6517	0,6310
	5	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6950	0,6503	0,6242
	10	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6950	0,6503	0,6242
	15	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6950	0,6503	0,6242
	20	0,5836	0,8275	0,7226	0,5746	0,5802	0,5667	0,6950	0,6503	0,6242
0,7	1	0,5836	0,5094	0,4935	0,3793	0,6686	0,5688	0,7030	0,6517	0,6310
	5	0,5836	0,8275	0,9847	0,7819	0,5802	0,5667	0,8317	0,7689	0,6854
	10	0,5836	0,8275	0,9847	0,7819	0,5802	0,5667	0,8317	0,7689	0,6854
	15	0,5836	0,8275	0,9847	0,7819	0,5802	0,5667	0,8317	0,7689	0,6854
	20	0,5836	0,8275	0,9847	0,7819	0,5802	0,5667	0,8317	0,7689	0,6854
0,8	1	0,5836	0,5094	0,4935	0,3793	0,6686	0,6007	0,6512	0,6057	0,6310
	5	0,5836	0,8275	0,8223	0,6483	0,5802	0,6133	0,6624	0,6191	0,7014
	10	0,5836	0,8275	0,8223	0,6483	0,5802	0,6133	0,6624	0,6191	0,7014
	15	0,5836	0,8275	0,8223	0,6483	0,5802	0,6133	0,6624	0,6191	0,7014
	20	0,5836	0,8275	0,8223	0,6483	0,5802	0,6133	0,6624	0,6191	0,7014
0,9	1	0,5836	0,5094	0,4935	0,3793	0,6686	1,1617	1,1396	0,6555	0,6687
	5	0,5836	0,8275	0,8223	0,6483	0,8517	0,6133	0,5744	0,5669	0,6818
	10	0,5836	0,8275	0,8223	0,6483	0,8226	0,6133	0,5744	0,5669	0,6818
	15	0,5836	0,8275	0,8223	0,6483	0,8226	0,6133	0,5744	0,5669	0,6818
	20	0,5836	0,8275	0,8223	0,6483	0,8226	0,6133	0,5744	0,5669	0,6818

Terlihat bahwa pada nilai DBI pada iterasi ke-15 dan iterasi ke-20 memiliki kesamaan nilai DBI pada iterasi ke-10, yang berarti nilai DBI sudah mulai stabil pada iterasi ke-10. Berdasarkan hal itu, jumlah iterasi yang digunakan pada metode *Self-Organizing Maps* ini sebanyak 10 iterasi.

4. Penentuan Jumlah Cluster Optimal dengan Davies Bouldin Index

Hasil analisis *cluster* atau *clustering* yang diperoleh menggunakan *Self-Organizing Maps* divalidasi untuk ditentukan jumlah *cluster* yang optimal menggunakan *Davies Bouldin Index*. Adapun nilai *Davies Bouldin Index* yang diperoleh ditunjukkan pada Tabel 6.

Tabel 6. DBI Hasil Clustering dengan Self-Organizing Maps

$k \backslash a$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
2	0,5836	0,5836	0,5836	0,5836	0,5836	0,5836	0,5836	0,5836	0,5836
3	0,5094	0,5094	0,8275	0,8275	0,8275	0,8275	0,8275	0,8275	0,8275
4	0,4935	0,7226	0,7226	0,8223	0,8223	0,7226	0,9847	0,8223	0,8223
5	0,3793	0,5746	0,5746	0,6483	0,6483	0,5746	0,7819	0,6483	0,6483
6	0,7409	0,5802	0,5802	0,5802	0,5802	0,5802	0,5802	0,5802	0,8226
7	0,5629	0,5667	0,5667	0,5667	0,5667	0,5667	0,5667	0,6133	0,6133
8	0,6167	0,6167	0,6187	0,7296	0,7296	0,6950	0,8317	0,6624	0,5744
9	0,5736	0,5736	0,5736	0,6859	0,6859	0,6503	0,7689	0,6191	0,5669
10	0,5867	0,5371	0,5867	0,5867	0,6242	0,6242	0,6854	0,7014	0,6818

Berdasarkan Tabel 6, dalam skenario jumlah *cluster* sebanyak 2 s.d. 10 *cluster*, diperoleh bahwa *Davies Bouldin Index* terkecil atau paling rendah yaitu 0,3793 yang terdapat pada jumlah *cluster* sebanyak 5. Oleh karena itu, hasil dari *clustering* ini diperoleh *cluster* yang

optimal berjumlah 5 cluster. Selanjutnya, dilakukan analisis karakteristik *cluster* terhadap jumlah *cluster* yang optimal.

5. Analisis Karakteristik *Cluster*

Setelah mendapatkan jumlah *cluster* yang optimal, dilanjutkan dengan pengelompokan provinsi di Indonesia berdasarkan indikator kriminalitas nasional. Anggota masing-masing *cluster* disajikan pada Tabel 7.

Tabel 7. Hasil *Clustering* Optimal yang Berjumlah 5 *Cluster*

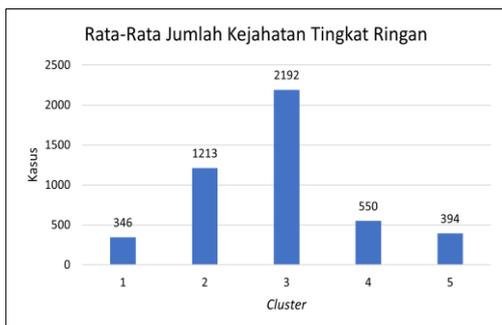
<i>Cluster</i>	Jumlah Anggota	Anggota <i>Cluster</i>
1	30	Aceh, Bali, Kep. Bangka Belitung, Banten, Bengkulu, Gorontalo, Papua Barat, Jambi, Jawa Barat, Jawa Tengah, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Lampung, Maluku Utara, Maluku, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Riau, Sulawesi Barat, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Barat, DI Yogyakarta, dan Sumatera Selatan
2	1	Sumatera Utara
3	1	Sulawesi Selatan
4	1	Jawa Timur
5	1	DKI Jakarta

Berdasarkan Tabel 7, diketahui bahwa masing-masing *cluster* memiliki beberapa anggota dengan rincian untuk *cluster* 1 terdiri dari 30 anggota, *cluster* 2 terdiri dari 1 anggota, *cluster* 3 terdiri dari 1 anggota, *cluster* 4 terdiri dari 1 anggota, dan *cluster* 5 terdiri dari 1 anggota. Setelah itu, akan dihitung rata-rata dari setiap variabel yang dimiliki oleh masing-masing *cluster* untuk mengetahui karakteristik dari *cluster* yang dibentuk. Hal ini disajikan pada Tabel 8.

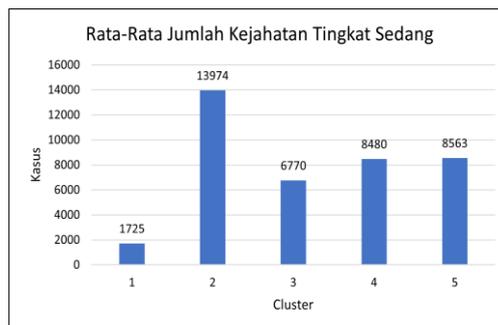
Tabel 8. Karakteristik *Cluster*

Variabel	Rata-rata				
	<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3	<i>Cluster</i> 4	<i>Cluster</i> 5
X1	346	1213	2192	550	394
X2	1725	13974	6770	8480	8563
X3	1640	14527	2012	6660	8651
X4	3283	24978	9405	10659	28518
X5	67,10	68,37	64,26	55,35	97,99
X6	10036	863	2154	1637	1083
X7	122	250	166	48	277

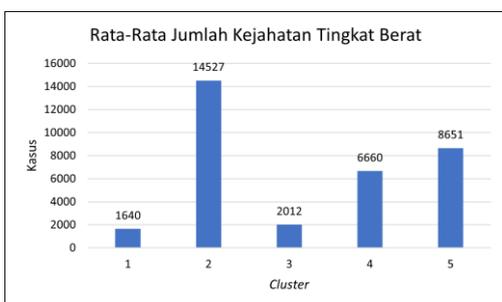
Adapun karakteristik *cluster* dapat direpresentasikan dalam bentuk diagram berdasarkan indikator kriminalitas nasional seperti pada Gambar 2 s.d. Gambar 8.



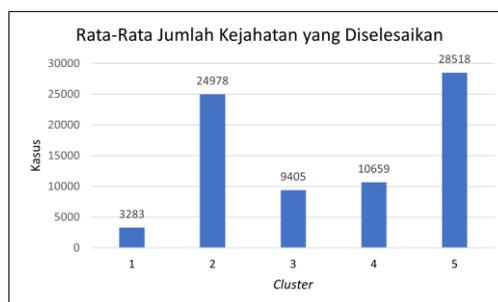
Gambar 2. Diagram Rata-Rata Jumlah Kejahatan Tingkat Ringan



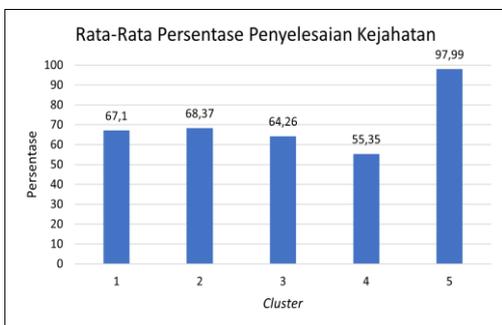
Gambar 3. Diagram Rata-Rata Jumlah Kejahatan Tingkat Sedang



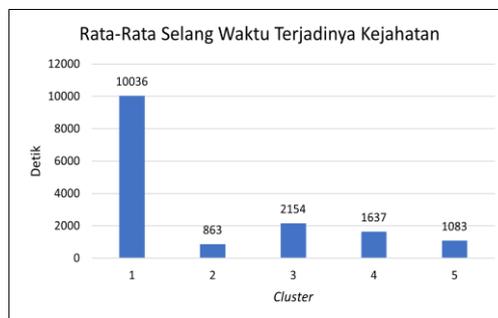
Gambar 4. Diagram Rata-Rata Jumlah Kejahatan Tingkat Berat



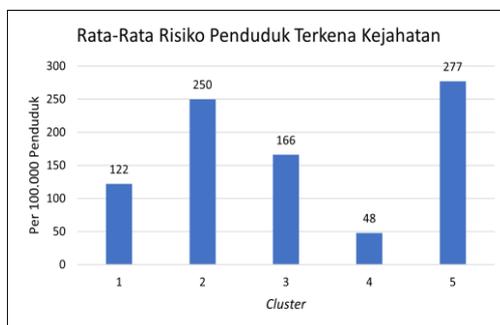
Gambar 5. Diagram Rata-Rata Jumlah Kejahatan yang Diselesaikan



Gambar 6. Diagram Rata-Rata Persentase Penyelesaian Kejahatan



Gambar 7. Diagram Rata-Rata Selang Waktu Terjadinya Kejahatan



Gambar 8. Diagram Rata-Rata Risiko Penduduk Terkena Kejahatan

Pembahasan

Hasil analisis *cluster* atau *clustering* dengan *Self-Organizing Maps* dan divalidasi dengan *Davies Bouldin Index* diperoleh bahwa jumlah *cluster* optimal adalah 5 *cluster* dengan *learning rate* 0,1. Adapun anggota dari setiap *cluster* optimal hasil analisis *cluster* dengan metode *Self-Organizing Maps* divisualisasikan pada Gambar 9.



Gambar 9. Pemetaan *Cluster*

Berdasarkan Gambar 9, terlihat bahwa mayoritas provinsi berwarna hijau. Sekelompok provinsi tersebut termasuk ke dalam *cluster* ke-1. Selain itu, provinsi yang berwarna merah termasuk ke dalam *cluster* ke-2, provinsi yang berwarna hijau muda termasuk ke dalam *cluster* ke-3, provinsi yang berwarna kuning termasuk ke dalam *cluster* ke-4, dan provinsi yang berwarna jingga termasuk ke dalam *cluster* ke-5. Setiap *cluster* memiliki karakteristiknya tersendiri. Adapun penjelasannya adalah sebagai berikut.

1. *Cluster* ke-1

Rata-rata jumlah kejahatan tingkat ringan, tingkat sedang, dan tingkat berat yang terjadi pada *cluster* ke-1 paling rendah dibandingkan *cluster* yang lain, begitupun dengan rata-rata jumlah kejahatan yang diselesaikan. Sementara, untuk rata-rata persentase penyelesaian kejahatan lebih tinggi daripada *cluster* ke-3 dan ke-4, tetapi lebih rendah daripada *cluster* ke-2 dan *cluster* ke-5. Kemudian, untuk rata-rata selang waktu terjadinya kejahatan pada *cluster* ke-1 ini memiliki waktu paling lama dibandingkan *cluster* yang lainnya sehingga diketahui bahwa rata-rata selang waktu terjadinya antar kejahatan adalah sekitar 167 menit. Selanjutnya, untuk rata-rata risiko penduduk terkena kejahatan berada di angka 122 yang mana lebih tinggi daripada *cluster* ke-4, tetapi lebih rendah daripada *cluster* ke-2, ke-3, dan ke-5. *Cluster* ini dapat disebut sebagai *cluster* yang paling aman di antara yang lain atau dapat digolongkan sebagai *cluster* dengan tingkat kriminalitas sangat rendah.

2. *Cluster* ke-2

Rata-rata jumlah kejahatan tingkat sedang dan tingkat berat yang terjadi pada *cluster* ke-2 paling tinggi dibandingkan *cluster* yang lain, tetapi rata-rata jumlah kejahatan tingkat ringan lebih rendah daripada *cluster* ke-3. Sementara, untuk rata-rata jumlah kejahatan yang diselesaikan dan rata-rata persentase penyelesaian kejahatan lebih rendah daripada *cluster* ke-5, tetapi lebih tinggi daripada *cluster* ke-1, ke-2, dan ke-4. Kemudian, untuk rata-rata selang waktu terjadinya kejahatan, pada *cluster* ke-2 ini memiliki waktu paling singkat dibandingkan *cluster* yang lainnya sehingga diketahui bahwa rata-rata selang waktu terjadinya antar kejahatan adalah sekitar 14 menit. Selanjutnya, untuk rata-rata risiko penduduk terkena kejahatan berada di angka 250 yang mana lebih rendah daripada *cluster* ke-5, tetapi lebih tinggi daripada *cluster* ke-1, ke-2, dan ke-4. *Cluster* ini dapat disebut sebagai *cluster* yang paling berbahaya di antara yang lain atau dapat digolongkan sebagai *cluster* dengan tingkat kriminalitas sangat tinggi.

3. Cluster ke-3

Rata-rata jumlah kejahatan tingkat sedang dan tingkat berat yang terjadi pada *cluster* ke-3 lebih tinggi daripada *cluster* ke-1, tetapi lebih rendah daripada *cluster* ke-2, ke-4, dan ke-5, begitupun dengan rata-rata jumlah kejahatan yang diselesaikan. Untuk rata-rata jumlah kejahatan tingkat ringan yang terjadi paling tinggi dibandingkan *cluster* yang lain. Sementara, untuk rata-rata persentase penyelesaian kejahatan lebih tinggi daripada *cluster* ke-4, tetapi lebih rendah daripada *cluster* ke-1, ke-2, dan ke-5. Kemudian, untuk rata-rata selang waktu terjadinya kejahatan, pada *cluster* ke-3 ini memiliki waktu yang lebih lama daripada *cluster* ke-2, ke-4, dan ke-5, tetapi lebih cepat daripada *cluster* ke-1 sehingga diketahui bahwa rata-rata selang waktu terjadinya antar kejahatan adalah sekitar 35 menit. Selanjutnya, untuk rata-rata risiko penduduk terkena kejahatan berada di angka 166 yang mana lebih rendah daripada *cluster* ke-2 dan ke-5 tetapi lebih tinggi daripada *cluster* ke-1 dan ke-4. *Cluster* ini dapat disebut sebagai *cluster* yang cukup aman dibandingkan *cluster* ke-2, ke-4, dan ke-5 atau dapat digolongkan sebagai *cluster* dengan tingkat kriminalitas rendah.

4. Cluster ke-4

Rata-rata jumlah kejahatan tingkat sedang dan tingkat berat yang terjadi pada *cluster* ke-4 lebih tinggi daripada *cluster* ke-1 dan ke-3, tetapi lebih rendah daripada *cluster* ke-2 dan ke-5, begitupun dengan jumlah kejahatan yang diselesaikan. Untuk rata-rata jumlah kejahatan tingkat ringan yang terjadi lebih tinggi daripada *cluster* ke-1 dan ke-5, tetapi lebih rendah daripada *cluster* ke-2 dan ke-3. Sementara, untuk rata-rata persentase penyelesaian kejahatan paling rendah dibandingkan *cluster* yang lain. Kemudian, untuk rata-rata selang waktu terjadinya kejahatan, pada *cluster* ke-4 ini memiliki waktu yang lebih lama daripada *cluster* ke-2 dan ke-5, tetapi lebih cepat daripada *cluster* ke-1 dan *cluster* ke-3 sehingga diketahui bahwa rata-rata selang waktu terjadinya antar kejahatan adalah sekitar 27 menit. Selanjutnya, untuk rata-rata risiko penduduk terkena kejahatan berada di angka 48 yang mana paling rendah dibandingkan *cluster* yang lain. *Cluster* ini dapat disebut sebagai *cluster* yang cukup berbahaya dibandingkan *cluster* ke-1 dan ke-3, tetapi cukup aman dibandingkan *cluster* ke-2 dan ke-5 atau dapat digolongkan sebagai *cluster* dengan tingkat kriminalitas sedang.

5. Cluster ke-5

Rata-rata jumlah kejahatan tingkat sedang dan tingkat berat yang terjadi pada *cluster* ke-5 lebih tinggi daripada *cluster* ke-1, ke-3, dan ke-4, tetapi lebih rendah daripada ke-2. Untuk rata-rata jumlah kejahatan tingkat ringan lebih tinggi daripada *cluster* ke-1, tetapi lebih rendah daripada *cluster* ke-2, ke-3, dan ke-4. Sementara, untuk rata-rata jumlah kejahatan yang diselesaikan dan persentase penyelesaian kejahatan paling tinggi dibandingkan *cluster* yang lain. Kemudian, untuk rata-rata selang waktu terjadinya kejahatan, pada *cluster* ke-5 ini memiliki waktu yang lebih lama daripada *cluster* ke-2, tetapi lebih cepat daripada *cluster* ke-1, ke-3, dan ke-4 sehingga diketahui bahwa rata-rata selang waktu terjadinya antar kejahatan adalah sekitar 18 menit. Selanjutnya, untuk rata-rata risiko penduduk terkena kejahatan berada di angka 277 yang mana paling tinggi dibandingkan *cluster* yang lain. *Cluster* ini dapat disebut sebagai *cluster* yang lebih aman dibandingkan *cluster* ke-2, tetapi lebih berbahaya dibandingkan *cluster* ke-1, ke-3, dan ke-4 atau dapat digolongkan sebagai *cluster* dengan tingkat kriminalitas tinggi.

SIMPULAN

Hasil analisis *cluster* dengan metode *Self-Organizing Maps* untuk pengelompokan provinsi di Indonesia berdasarkan indikator kriminalitas nasional menunjukkan bahwa jumlah *cluster* sebanyak 5 terpilih menjadi jumlah *cluster* optimal dengan nilai DBI 0,3793 dari pengujian skenario jumlah *cluster* dari 2 s.d. 10 *cluster*. *Cluster* ke-1 beranggotakan 30 provinsi, *cluster* ke-2 beranggotakan 1 provinsi, *cluster* ke-3 beranggotakan 1 provinsi, *cluster* ke-4 beranggotakan 1 provinsi, dan *cluster* ke-5 beranggotakan 1 provinsi.

Karakteristik dari hasil analisis *cluster* dengan metode *Self-Organizing Maps* untuk pengelompokan provinsi di Indonesia berdasarkan indikator kriminalitas nasional ditunjukkan dengan rincian berikut. *Cluster* ke-1 memiliki rata-rata jumlah kejahatan tingkat ringan, tingkat sedang, tingkat berat, dan rata-rata jumlah kejahatan yang diselesaikan paling rendah, serta memiliki rata-rata selang waktu terjadinya kejahatan paling lama. *Cluster* ke-2 memiliki rata-rata jumlah kejahatan tingkat sedang dan tingkat berat paling tinggi, serta memiliki rata-rata selang waktu terjadinya kejahatan paling cepat. *Cluster* ke-3 memiliki rata-rata jumlah kejahatan tingkat ringan paling tinggi. *Cluster* ke-4 memiliki rata-rata persentase penyelesaian kejahatan dan rata-rata risiko penduduk terkena kejahatan paling rendah. *Cluster* ke-5 memiliki rata-rata jumlah kejahatan yang diselesaikan, rata-rata persentase penyelesaian kejahatan, dan rata-rata risiko penduduk terkena kejahatan paling tinggi. Secara berturut-turut dari *cluster* ke-1 s.d. *cluster* ke-5 dapat digolongkan sebagai *cluster* dengan tingkat kriminalitas sangat rendah, sangat tinggi, rendah, sedang, dan tinggi.

Hasil penelitian ini diharapkan dapat menjadi evaluasi bagi pihak pemerintah sehingga dapat menetapkan suatu kebijakan yang lebih tegas lagi terkait adanya kasus kriminalitas atau kejahatan di Indonesia agar masyarakat merasa aman dalam menjalani kehidupan. Di sisi lain, masyarakat juga dapat turut andil dalam menjaga keamanan di mana pun dan kapan pun karena ini merupakan tugas bersama antara pihak pemerintah dan masyarakat. Untuk penelitian selanjutnya, dapat menggunakan variabel yang lainnya yang berkaitan dengan indikator kriminalitas nasional dan juga dapat menggunakan metode analisis *cluster* yang lainnya.

UCAPAN TERIMAKASIH

Terimakasih kepada koor Prodi Matematika dan seluruh Dosen Matematika yang telah memberikan ilmu serta bimbingan hingga terselesainya artikel ini.

DAFTAR PUSTAKA

- Arthur, D. & Vassilvitskii, S. (2007). K-Means++ : The Advantages of Careful Seeding. Badan Pusat Statistik. (2021). Statistik Kriminal 2021. Jakarta : Badan Pusat Statistik. Badan Pusat Statistik. (2022). Statistik Kriminal 2022. Jakarta : Badan Pusat Statistik. Bustamam, A., Rivai, M. A., & Siswantining, T. (2018). Implementation of Spectral Clustering on Microarray Data of Carcinoma Using Self Organizing Map (SOM). <https://doi.org/10.1063/1.5064237>
- Choudary, A. (2019). Comprehensive Guide to K-Means Clustering. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Dulkiah, M. (2020). *Sosiologi kriminal*. http://digilib.uinsgd.ac.id/40086/1/BUKU_SOSIOLOGI_KRIMINAL.pdf
- Gupta, N. (2021). ML | K-Means Algorithm. GeeksforGeeks. <https://www.geeksforgeeks.org/ml-k-means-algorithm/>
- Halim, N. N., & Widodo, E. (2017). Clustering dampak gempa bumi di indonesia menggunakan kohonen self organizing maps. *Prosiding SI MaNIS (Seminar Nasional Integrasi Matematika Dan Nilai Islami)*, 1(1), 188–194. <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62>
- Khotija. (2022). Pengelompokan Tindak Kriminal di Indonesia Menggunakan Algoritma Self Organizing Maps (SOM) dan K-Means.
- Lestari. (2018). Pengelompokan Provinsi Di Indonesia Berdasarkan Kriminalitas

- Menggunakan Metode Ward Dan K-Medoids. *Repository.Unimus.Ac.Id*.
[http://repository.unimus.ac.id/4348/12/19.Jurnal Ilmiah_ANGGI.pdf](http://repository.unimus.ac.id/4348/12/19.Jurnal%20Ilmiah_ANGGI.pdf)
- Numbeo. (2022). South-Eastern Asia: Crime Index by Country 2022 Mid-Year.
https://www.numbeo.com/crime/rankings_by_country.jsp?title=2022-mid®ion=035&displayColumn=0.
- P., M. A., Hendrawan, R. A., & Hafidz, I. (2014). K-Means Untuk Klasifikasi Pelanggan Perusahaan. *Seminar Nasional Sistem Informasi Indonesia (SESINDO)*, September, 419–425.
- Pamulang, M. N. P., Aini, M. N., & Enri3, U. (2021). Komparasi Distance Measure Pada K-Medoids Clustering untuk Pengelompokan Penyakit ISPA. *Edumatic: Jurnal Pendidikan Informatika*, 5(1), 99–107. <https://doi.org/10.29408/edumatic.v5i1.3359>
- Septiani, I. W., Fauzan, A. C., & Huda, M. M. (2022). Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 556. <https://doi.org/10.30865/json.v3i4.4055>
- Tempola, F., Muhammad, M., & Mubarak, A. (2020). Penggunaan Internet Dikalangan Siswa SD di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering dan Validasi DBI. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(6), 1153. <https://doi.org/10.25126/jtiik.2020722370>
- Undang-Undang Nomor 1 Tahun 2023 Tentang Kitab Undang-Undang Hukum Pidana.
- Wahyudi, N., Suhartono, V., & Pramunendar, R. A. (2015). Background Subtraction Berbasis Self Organizing Map Untuk Deteksi Objek Bergerak. *Systemic: Information System and Informatics Journal*, 1(1), 42–51. <https://doi.org/10.29080/systemic.v1i1.283>
- Wehrens, R. (2007). <Kohonen-Manual.Pdf>. *JSS Journal of Statistical Software*, 21(5). <http://www.jstatsoft.org/>
- Yang, L., Ouyang, Z., & Shi, Y. (2012). A modified clustering method based on self-organizing maps and its applications. *Procedia Computer Science*, 9, 1371–1379. <https://doi.org/10.1016/j.procs.2012.04.151>