UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 3 Nomor 1 Tahun 2025 Halaman 39-46.
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Implementation of Self-Organizing Maps (SOMs) Method in Clustering Indonesian Provinces Based on the Number of Crimes by Type of Crime

Fajriyanti Nur Putri, Tessy Octavia Mukhti*, Nonong Amalita dan Admi Salma
Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia
*Corresponding author: tessyoctaviam@fmipa.unp.ac.id

Submitted : 20 Januari 2025
Revised : 12 Februari 2025
Accepted : 28 Februari 2025

ABSTRACT

Crime cases are often the main topic of daily news in various media in Indonesia. Some of these crime cases are detrimental to the surrounding community and some are detrimental and these actions cannot be avoided in human life because they have become one type of social phenomenon. To protect the community by providing a sense of security and peace, the Indonesian government, especially the police, must pay attention to conditions like this. The results of this study used the Self Organizing Maps (SOMs) method to obtain 3 clusters with the characteristics of each cIuster. The first cluster with a low impact crime rate consists of 29 provinces. The second cluster with a moderate impact consists of 3 provinces showing the most dominant crime rate, namely crimes related to fraud, embezzlement, smuggling & corruption compared to other clusters. The third cluster with a high impact consists of 2 provinces with the most prominent characteristics by showing almost all indicators of the number of crimes according to the type of crime experiencing the highest average crime cases compared to other clusters.

Keywords: *Clusters, Crime, SeIf Organizing Maps (SOMs)*

I. PENDAHULUAN

Data *mining* ada1ah proses untuk menemukan pola-pola yang tersembunyi dengan memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembe1ajaran mesin guna mengekstrak dan mengidentifikasi informasi yang berpotensi berguna yang ada dalam basis data besar. (Wicaksono, 2016). Proses ini memungkinkan pengumpulan dan pengolahan data untuk mengungkap informasi penting dari data (Han, Pei, and Tong 2022). Berbagai teknik dalam data *mining* digunakan, seperti *clustering, classification, prediction, association rules,* dan lainnya (Khormarudin 2016). *Clustering* adalah proses pembelajaran untuk mengelompokkan dataset ke dalam subkelompok yang diinginkan, di mana setiap objek dalam kelompok memiliki kesamaan berdasarkan matriks tertentu (Talakua, Leleury, and Taluta 2017).

Dalam data *mining*, terdapat dua jenis metode *c1ustering*, yaitu *hierarchicaI* dan *non-hierarchical* (Khormarudin 2016). Metode *hierarchical* digunakan untuk mengelompokkan dua objek atau lebih dengan nilai kemiripan terdekat (Muhidin 2017). Sedangkan *metode non-hierarchical* digunakan untuk mengelompokkan data dengan jumlah kelompok yang dapat ditentukan sebelumnya (Alamtaha, Djakaria, and Yahya 2023).

Salah satu metode *clustering* dalam data *mining* adalah *Self Organizing Maps (SOMs),* yang diperkenalkan oleh Profesor Teuvo Kohonen pada tahun 1982. Metode ini menggunakan neuron pemenang yang memiliki bobot untuk diperbarui dan tidak bergantung pada nilai target kelas, sehingga tidak ada kelas yang ditugaskan untuk setiap data (Kohonen and Kohonen 1989). Selama proses pengaturan diri, kelompok dengan vektor bobot yang paling cocok dengan pola input (dengan jarak terdekat) akan menjadi pemenang, dan neuron yang menang serta tetangganya akan menyesuaikan bobot mereka (Harun 2015).

SOM merupakan metode pelatihan tanpa pengawasan (*unsupervised learning)*, yang digunakan untuk mengelompokkan data berdasarkan karakteristik atau fitur-fitur data. Ini membentuk topologi dari *Unsupervised ArtificiaI NeuraI Network (Unsupervised ANN)* (Shieh and Liao 2012). Memungkinkan visualisasi dan proyeksi data berdimensi tinggi ke dalam dimensi rendah, sehingga memudahkan pemahaman struktur data dan pengelompokan data. Salah satu karakteristik SOM adalah kemampuannya untuk memvisualisasikan hasil pengelompokan dalam bentuk topografi dua dimensi, seperti peta, yang memudahkan pengamatan distribusi kelompok hasil pengelompokan.

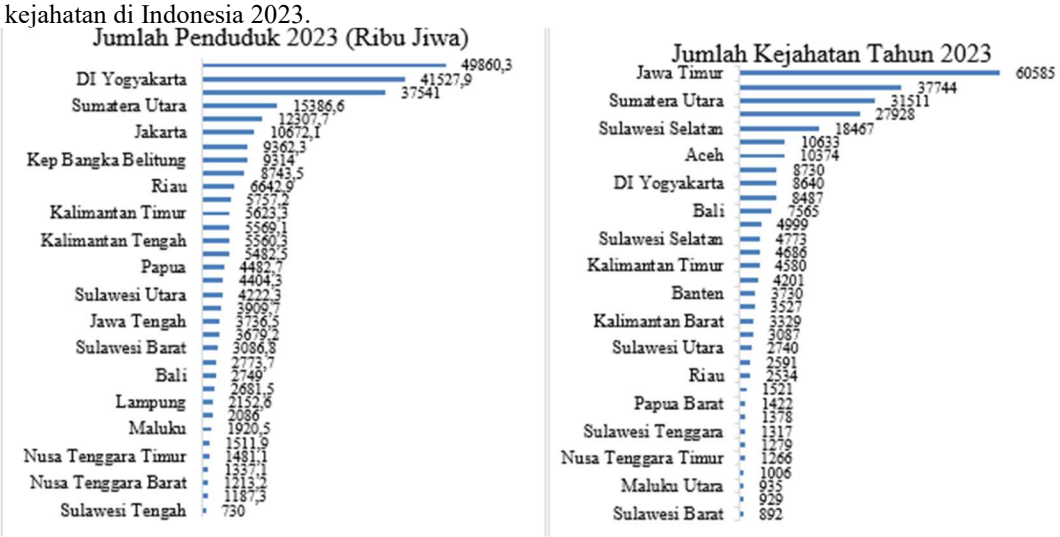
DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Penggunaan metode SOM memi1iki akurasi yang 1ebih baik dibandingkan metode *clustering* lainnya. Penelitian yang dilakukan oleh (Fawaz 2023) dengan judul Perbandingan Algoritma *Self Organizing Map* dan *Fuzzy C-Means* dalam *Clustering* Hasil Produksi Ikan PPN Karangantu menunjukkan, bahwa metode SOM memiliki akurasi lebih tinggi dibandingkan dengan *Fuzzy C-Means* (FCM). Pada penelitian ini memperoleh bahwa metode SOMs memi1iki akurasi lebih baik pada pengelompokan FCM, metode SOM juga menunjukkan ni1ai *error* terkecil. Ha1 ini juga dibuktikan oleh (Munawar and Kunci 2015) dalam penelitiannya Implementasi Algoritma *Self Organizing Map* (SOM) untuk *Clustering* Mahasiswa pada Mata kuliah Proyek (Studi Kasus: JTK POLBAN), yang menyatakan bahwa semakin kecil nilai mean square error (MSE), maka semakin konvergen *cluster* yang terbentuk.

Berdasarkan pemaparan tersebut, maka penulis tertarik untuk mengkaji lebih lanjut mengenai metode SOMs yang memberikan akurasi yang baik, yaitu hasil *cluster* yang mudah ditafsirkan atau dipahami melalui *map feature* yang memiliki bentuk persegi atau heksagonal. Metode SOMs ini dapat menghasilkan data non-linear secara inheren, artinya dapat menangkap hubungan kompleks antar variabel. Dengan menggunakan metode SOMs ini dapat mengelompokan provinsi di Indonesia berdasarkan jumlah kejahatan menurut jenis kejahatan.

Kejahatan merupakan tindakan yang melanggar hukum yang dapat menyebabkan seseorang dihukum. Tindak kejahatan dapat terjadi pada siapa saja, tanpa memandang latar belakang dan hal ini telah berlangsung sepanjang zaman. Kasus kejahatan ini ada yang merugikan masyarakat sekitar dan ada juga yang merugikan diri sendiri (Simatupang and Faisal 2017). Kejahatan di Indonesia perlu di kelompokkan untuk melihat karakteristik jumlah kejahatan menurut jenis kejahatan, karena tingkat kejahatan di Indonesia relatif tinggi bahkan menempati peringkat 20 dalam tingkat kejahatan paling tinggi di dunia. Penyebab tingginya tindak kejahatan di Indonesia dapat dikaitkan dengan faktor jumlah penduduk yang terus berkembang pesat. Peningkatan jumlah penduduk sering kali beriringan dengan meningkatnya tekanan sosial dan ekonomi, seperti kemiskinan, pengangguran, dan ketidaksetaraan, yang dapat mendorong individu untuk terlibat dalam tindakan kriminal. Selain itu, padatnya populasi juga dapat memperburuk masalah urbanisasi, menyebabkan ketegangan sosial dan mempersempit akses terhadap layanan dasar, yang pada akhirnya mempengaruhi tingkat kejahatan. Berikut data Badan Pusat Statistik (BPS) mengenai jumlah penduduk dan jumlah kejahatan di Indonesia 2023.



Gambar 1. Jumlah Penduduk dan Jumlah Kejahatan Tahun 2023

Berdasarkan Gambar 1 dapat dilihat bahwa provinsi di Indonesia yang memiliki jumlah kejahatan yang bervariasi, hal ini diakibatkan karena perbedaan karakteristik jenis kejahatan dari setiap provinsi di Indonesia. Namun perbedaan banyaknya kejadian kejahatan ini tidak selalu berbanding lurus dengan jumlah penduduk yang ada di Indonesia. Provinsi Jawa Timur yang memiliki jumlah kejahatan yang paling tinggi di Indonesia pada tahun 2023 sebanyak 60585 kasus, tapi jumlah penduduknya untuk tahun 2023 sebanyak 12.307,7 ribu jiwa di provinsi Jawa Timur tidak sebanyak 49.860,3 ribu jiwa di provinsi Jawa Barat pada tahun 2023.

Jumlah penduduk di Indonesia yang tinggi tidak menjadi salah satu faktor utama banyaknya terjadi kejahatan. Walaupun provinsi Jawa Barat yang meiliki jumlah penduduk yang pa1ing tinggi di lndonesia, tapi tingkat kejahatan yang terjadi berada di urutan ke 4 yang tinggi yaitu sebanyak 27928 kasus. Selain itu Provinsi Sulawesi Tengah

DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

memliki jumlah penduduk sebanyak 730 ribu jiwa paling sedikit daripada provinsi Sulawesi Barat yang sebanyak 3.086,8 ribu jiwa, tapi tingkat kejahatan di provinsi Sulawesi Barat paling sedikit yaitu 892 kasus.

Berdasarkan permasalahan ini, penelitian ini bertujuan untuk membahas "Implementasi Metode *Self Organizing Maps* (SOMs) dalam Pengelompokan Provinsi di lndonesia Berdasarkan Jumlah Kejahatan Menurut Jenis Kejahatan".

II. METODE PENELITIAN

Pada penelitian ini menggunakan data sekunder yang dipero1eh dari website Badan Pusat Statistik (BPS) yaitu buku publikasi Statistik Kriminal 2024. Berdasarkan dari buku tersebut, data jumlah kejahatan untuk seluruh di Indonesia akan dikelompokkan berdasarkan menurut jenis kejahatannya. Terdapat sembilan variabel yang akan digunakan, yaitu: kejahatan terhadap nyawa (X1), kejahatan terhadap fisik (X2), kejahatan terhadap kesusi1aan (X3), kejahatan terhadap kemerdekaan orang (X4), kejahatan terhadap hak mi1ik/barang dengan penggunaan kekerasan (X5), kejahatan terhadap hak mi1ik/barang (X6), kejahatan terkait narkotika (X7), kejahatan terkait pemipuan, penggelapan, & korupsi (X8), dan kejahatan terhadap ketertiban umum (X9). Langkah-langkah yang akan digunakan dalam metode SOMs sebagai berikut:

1. Melakukan Standardisasi data.
   Standardisasi dilakukan jika proses perhitungan suatu analisis yang tidak valid salah satu penyebab adalah terdapat pencilan (*outlier*) yang mana perbedaan satuan yang besar dari variabel-variabel yang diteliti. Menurut (Simamora, 2005) Cara menentukan nilai standardisasi adalah dengan menggunakan persamaan berikut:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

   Dimana,
   z_{ij} : Data hasil standardisasi observasi ke-*i* variabe1 ke-*j*
   x_{ij} : Observasi ke-*i* variabel ke-*j*
   $\bar{x}_j$: Rata-rata variabel ke-*j*
   S_j : Simpangan baku variabel ke-*j*
2. Menentukan jumlah *cluster* dengan menggunakan metode *Elbow.*
   Langkah pertama untuk melakukan analisis *cluster* adalah menentukan banyak *cluster* k terbaik dari tiga pendekatan yang bisa digunakan dalam menentukan jumlah k terbaik ialah metode *Elbow*, *Silhouette Coefficient*, dan *Gap Statistic*. Menurut (Dewi & Pramita, 2019) untuk mendapatkan perbandingannya ialah menghitung *Sum of Square Error* (*SSE*) dari masing-masing nilai *cluster*. Karena semakin besar dari jumlah nilai *cluster* k, maka untuk nilai *SSE* akan semakin kecil. Berikut rumus SSE didefinisikan:

$$SSE = \sum_{K=1}^{K} \sum_{x_i} |x_i - c_k|^2 \quad (2)$$

   Dimana,
   K : *cluster* ke-*c*
   x_i : jarak data objek ke-*i*
   c_k : pusat *cluster* ke-*k*
3. Menentukan jenis dan ukuran topologi (*x*, *y*), *learning rate* (α), dan ukuran ketetanggaan (R).
4. Menginisialisasikan bobot neuron *output* secara acak dengan rentang 0 < w < 1.
5. Menghitung jarak minimum. Data *input* akan dipilih secara acak kemudian akan dihitung jaraknya dengan masing-masing neuron *output* sampai menemukan *wining* neuron (neuron dengan jarak terkecil). Menghitung jarak tersebut dengan mengunakan Jarak *Euclidean* dengan persamaan yaitu:

$$D_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \quad (3)$$

   Keterangan:
   D_{ij} : Jarak Euclidean antara data *input* pada baris ke-*i* dan neuron ke-*j*
   x_{ik} : nilai amatan antara objek ke-*i* dan objek ke-*k*
   x_{jk} : nilai amatan antara objek ke-*j* dan objek ke-*k*
   n : banyak variabel
6. Memperbarui nilai bobot neuron *output* yang menjadi *winning* neuron dengan menggunakan persamaan berikut:

$$w_{ij,new} = w_{ij,old} + \alpha(x_{ij} - w_{ij,old}) \quad (4)$$

DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Dimana,
$w_{ij,new}$: bobot baru dari neuron baris, baris ke *i* dan kolom ke *j*
$w_{ij,old}$: bobot lama dari neuron baris, baris ke *i* dan kolom ke *j*
α : *learning rate*
x_{ij} : data input pada baris ke *i* dan kolom ke *j*
Setelah itu, memperbarui semua neuron tetangga menggunakan persamaan sebagai berikut:

$$w_{ij,new} = w_{ij,old} + \alpha * h_{ci}(x_{ij} - w_{ij,old}) \quad (5)$$

Dimana,
h_{ci} : nilai fungsi untuk neuron pemenang c dengan neuron tetangga ke *i*
Nilai fungsi h_{ci} dapat dihitung dengan menggunakan persamaan sebagai berikut:

$$h_{ci} = \exp\left(\frac{d_{ci}^2}{2R^2}\right) \quad (6)$$

Dimana,
d_{ci} : jarak antara neuron pemenang c dengan neuron tetangga ke *i*
R : ukuran ketetanggaan

7. Melakukan langkah 4-6 untuk semua data *input*.
8. Memeriksa hasil *cluster* dengan memanfaatkan validasi internal.
   Menurut Jain & Dubes (1988:161) validasi *cluster* ialah prosedur yang mengevaluasi hasil analisis *cluster* secara kuantitatif dan objektif. Dan terdapat tiga pendekatan untuk mengeksplorasi validitas *cluster*:

a. *Connectivity*
   Pada Indeks *Connectivity* mempunyai nilai nol sampai ∞, jika nilai koefisien *connectivity* dinyatakan baik apabila nilainya semakin rendah pada *cluster* yang terbentuk (Brock, G., dkk, 2008). Rumus Indeks *Connectivity* didefinisikan sebagai berikut:

$$\boldsymbol{Conn(C)} = \sum_{i=1}^{N}\sum_{j=1}^{L} \boldsymbol{x_{i,nn_{i(j)}}} \quad (7)$$

   Dimana,
   $\boldsymbol{nn_{i(j)}}$ = tetangga terdekat dari objek *j* ke objek di-*i*
   L = banyak *c1uster*
   N = banyak pengamatan
   $\boldsymbol{x_{i,nn_{i(j)}}}$ = nilai pada objek ke- *i* bernilai 0 jika objek *i* dan *j* dalam satu *cluster* dan nilai 1/j Ketika sebaliknya

b. *Silhouette*
   Silhouette merupakan salah satu ukuran validasi yang berbasis kriteria internal. Nilai *silhouette* yang digunakan ada1ah -1≤ $\boldsymbol{s_i}$ ≤1. Ketika *silhouette* akan mengevaluasi untuk penempatan setiap objek dalam setiap *cluster* dengan membandingkan jarak rata-rata objek dalam satu *cluster* dan jarak antara objek dengan *cIuster* yang berbeda (Brock, G., dkk, 2008). Rumus indeks *Silhouette* didefinisikan sebagai berikut:

$$\boldsymbol{S_{(i)}} = \frac{\boldsymbol{b_{(i)} - a_{(i)}}}{\boldsymbol{\max(a_{(i)}, b_{(i)})}} \quad (8)$$

   Dimana,
   $\boldsymbol{S_{(i)}}$: *silhouette* pada titik ke- *i.*
   $\boldsymbol{a_{(i)}}$: rata-rata kemiripan antara objek ke- *i* dengan objek lain di dalam *cluster.*
   $\boldsymbol{b_{(i)}}$: nilai minimum dari rata-rata kemiripan antara objek ke- *i* dengan objek lain di luar *cluster* nya.

c. Indeks *Dunn*
   Menurut Velelz & Ortega (2002) lndeks ini mencoba untuk mengidentifikasi kompak dan terpisah dengan baik *cluster*. lndeks Dunn merupakan perbandingan antara jarak minimum antar observasi pada *cluster* yang berbeda dengan jarak maksimum yang terdapat di dalam satu *cluster*. lndeks Dunn dihitung menggunakan rumus berikut:

$$\boldsymbol{Dunn} = \frac{\boldsymbol{d_{min}}}{\boldsymbol{d_{max}}} \quad (9)$$

DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Dimana,

d_{min} : jarak terkecil antara observasi pada *cluster* yang berbeda

d_{max} : jarak terbesar pada masing-masing *cluster* data

Nilai indeks yang besar menunjukkan adanya kompak dan *cluster* yang terpisah dengan baik.

9. Melakukan interpretasi dari *cluster* yang diperoleh.

III. HASIL DAN PEMBAHASAN

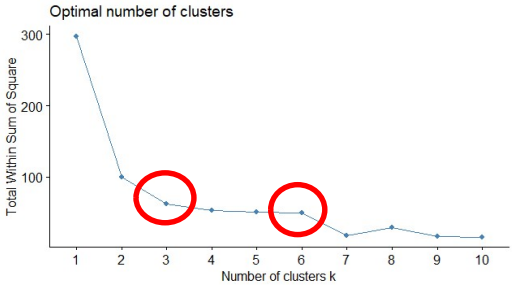
Langkah awal dalam analisis *cluster* menggunakan algoritma SOMs adalah melakukan proses standardisasi data. Standardisasi dilakukan untuk mengatasi perbedaan skala data yang signifikan, sehingga setiap variabel memiliki kontribusi yang setara dalam analisis. Hasil dari proses standardisasi ini disajikan pada Tabel 1.

Tabel 1. Data Hasil Standardisasi

	1	2	3	4	:	9
1	-0,34	0,17	0,45	-0,13	..	-1,26
2	1,61	2,65	2,3	2,95	..	8,02
3	-0,18	0,33	0,08	0,04	..	-7,69
4	-0,4	-0,57	-0,4	-0,54	..	-3,92
5	-0,22	-0,32	-0,36	-0,006	..	-3,92
:	:	:	:	:	:	:
34	-0,36	-0,59	-0,48	-0,66	..	-4,07

a. Menentukan Jumlah *Cluster*

Langkah pertama dalam melakukan analisis *cluster* adalah menentukan jumlah *cluster*. Pada penelitian ini, jumlah *cluster* ditentukan menggunakan metode *elbow* dengan melihat nilai *Within Sum of Square* (WSS).

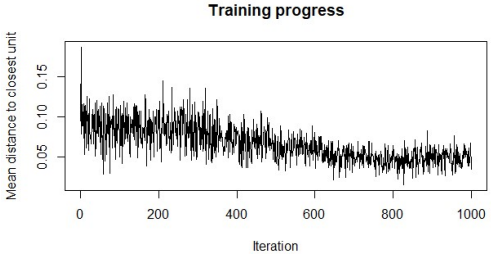


Gambar 2. *Within Sum of Square*

Berdasarkan Gambar 2, terlihat bahwa ketika menggunakan metode *elbow*, nilai *Within Sum of Squares* menunjukkan perubahan yang mulai landai pada indeks *cluster* 3 hingga 6. Hal ini berbeda dengan perubahan pada indeks sebelumnya yang terlihat cukup curam. Berdasarkan hasil metode elbow tersebut, peneliti memutuskan untuk menentukan jumlah *cluster* sebanyak 3 kelas.

b. Penerapan metode (SOMs)

Analisis SOMs perlu melakukan suatu *training progress* untuk meminimalisir rata-rata jarak suatu objek ke unit terdekat (Wehrens dan Buydens, 2007).



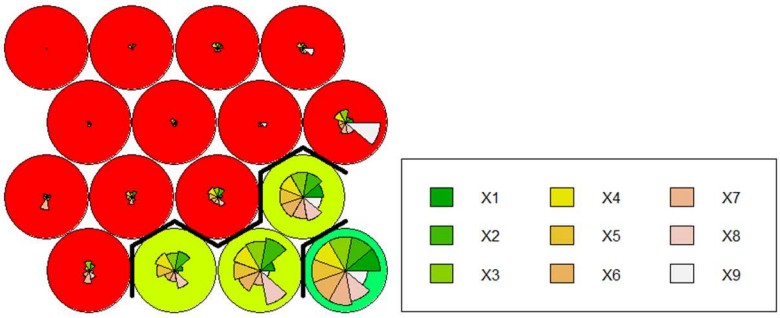
Gambar 3. *Output Training Progress*

DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Gambar 3 menjelaskan jumlah *training progress* yang menunjukkan banyaknya iterasi yang dilakukan untuk mengoptimalkan jarak hingga mendekati rata-rata terdekat, sampai hasilnya mencapai kondisi konvergen. Iterasi konvergen adalah proses yang dilakukan oleh *software* hingga menghasilkan hasil yang stabil. Dalam hal ini, iterasi mulai konvergen pada iterasi ke-500 dan berakhir pada iterasi ke-1000. Semakin banyak iterasi yang dilakukan, nilai *mean distance to closest unit* menjadi semakin kecil, yang berdampak pada kua1itas *cluster* yang dihasilkan menjadi semakin baik. Nilai *mean distance to closest unit* mulai stabil ketika berada di bawah 0,05.

Selanjutnya, dalam proses analisis SOMs ini, diperoleh model SOMs yang menghasilkan diagram kipas menggunakan *software* R-Studio, seperti yang ditampilkan pada gambar di bawah ini.



Gambar 4. Diagram Kipas (*fan*) tentang *cluster* yang terbentuk di masing-masing *cluster*

Dalam analisis SOMs ini, peneliti menggunakan *Hexagonal Topology* karena topologi ini memungkinkan distribusi yang merata pada grid 4x4 dalam visualisasi *cluster*. Pada diagram kipas yang digunakan dalam penelitian ini, distribusi variabel divisualisasikan, dimana semakin besar ukuran kipas yang terbentuk, semakin tinggi nilai variabel tersebut. Diagram kipas ini terdiri dari neuron-neuron berbentuk lingkaran yang mencerminkan karakteristik anggota di setiap neuron. Neuron-neuron yang terbentuk kemudian dikelompokkan dengan neuron lain yang memiliki tingkat kemiripan paling tinggi.

Pada Gambar 4, terlihat adanya 3 *cluster* yang ditandai dengan warna berbeda, yaitu merah, kuning, dan hijau. Perbedaan warna ini merepresentasikan kondisi masing-masing *cluster* berdasarkan variabel yang terkait dengan jumlah kejahatan menurut jenis kejahatan tahun 2023. Batas antar *cluster* divisualisasikan dengan garis hitam pada diagram kipas. Untuk informasi lebih rinci mengenai hasil *cluster* dan karakteristik masing-masing *cluster*, dapat merujuk pada Tabel 2, yang memaparkan hasil analisis *cluster* berdasarkan jumlah kejahatan menurut jenis kejahatan di setiap provinsi di Indonesia menggunakan *software* R-Studio

Tabel 2. Anggota *Cluster* yang Terbentuk

Cluster	**Jumlah Anggota *Cluster***	**Anggota *Cluster***
1	29	Aceh, Sumatera Barat, Riau, Bengkulu, Jambi, Sumatera Selatan, Kep. Bangka Belitung, Kepulauan Riau, Jakarta, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua, Papua Barat, Sulawesi Barat, dan Kalimantan Tenggara.
2	3	Lampung, Jawa Barat, dan Sulawesi Selatan.
3	2	Sumatera Utara dan Jawa T1mur

Berdasarkan tabe1 2, diketahui bahwa *CIuster* 1 terdiri dari 29 anggota yang ditandai dengan lingkaran berwarna merah. *CIuster* 2 memiliki 3 anggota yang ditandai dengan lingkaran berwarna kuning, sedangkan *cluster* 3 terdiri dari 2 anggota yang ditandai dengan lingkaran berwarna hijau.

c. Profilisasi *cluster*

Setelah menentukan jumlah anggota di setiap *cluster*, langkah selanjutnya adalah menghitung rata-rata setiap *cIuster* untuk mengidentifikasi karakteristik masing-masing *cIuster* (profilisasi) berdasarkan indikator

DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

jumlah kejahatan menurut jenis kejahatan di Provinsi Indonesia tahun 2023. Sebelum proses profilisasi dilakukan, data yang digunakan adalah data mentah yang belum distandardisasi. Tabel 3 berikut menunjukkan hasil perhitungan profilisasi *cluster*.

Tabel 3. Profilisasi *cluster*

Variabe1	**Rata-rata Profilisasi *Cluster***		
	Cluster 1	*Cluster* 2	*Cluster* 3
X1	515	241	523
X2	16586	10965	10051
X3	4236	2689	2724
X4	3710	1962	1860
X5	901	891	939
X6	22876	14256	21014
X7	16342	4592	8736
X8	9840	7795	6874
X9	64098	13637	39375

Tabel 3 memperlihatkan profilisasi atau karakteristik masing-masing *cluster* yang terbentuk. Jika suatu *cluster* memiliki nilai rata-rata tertinggi, maka karakteristik tersebut menjadi yang paling dominan dibandingkan *cluster* lainnya. *Cluste*r 1 menunjukkan dominasi tertinggi dalam hal jumlah kejahatan berdasarkan jenis kejahatan di Provinsi Indonesia tahun 2023. *Cluster* 2 lebih dominan pada variabel Kejahatan terhadap Kesusilaan (X_3), Kejahatan terhadap Hak Milik/Barang dengan Penggunaan Kekerasan (X_5), Kejahatan terhadap Hak Milik/Barang (X_6), serta Kejahatan terkait Penipuan, Penggelapan, Penyelundupan, dan Korupsi (X_8). Sementara itu, *Cluster 3* memiliki karakteristik yang kurang dominan dibandingkan *cluster* lainnya, tetapi lebih menonjol pada variabel Kejahatan terhadap Nyawa (X_1), Kejahatan terkait Narkotika (X_7), dan Kejahatan terhadap Ketertiban Umum (X_9).

d. Validasi *Cluster*

Selanjutnya, melakukan uji validasi untuk menilai apakah jumlah *cluster* yang diperoleh dari metode Kohonen SOMs sudah tepat. Uji validasi ini menggunakan pendekatan internal untuk mengevaluasi kualitas pembagian data menjadi beberapa kelompok dengan memanfaatkan validasi internal.

Tabel 4. *Output* Validasi *Cluster* dari *Software* R

Metode	Ukuran *Cluster*				
	3	4	5	6	7
Dunn	0,55	0,11	0,07	0,06	0,06
Silhouette	0,65	0,41	0,30	0,27	0,23
Connectivity	12,96	22,32	30,34	27,44	36,22

Tabel 4 menunjukkan bahwa pembagian data menjadi 3 cluster menghasilkan hasil validasi internal terbaik. Nilai indeks *Connectivity* yang rendah pada *cluster* 3 dengan nilai 12,96, mengindikasikan bahwa objek dalam satu *cluster* memiliki hubungan yang kuat satu sama lain. Selain itu, nilai indeks *Dunn* yang mendekati 1 dan pada ukuran *cluster* 3 dengan nilai 0,55, sedangkan nilai *Silhouette* yang tinggi pada cluster 3 dengan nilai 0,65 menunjukkan bahwa pemisahan antar *cluster* sangat baik. Maka dari itu hasil validasi internal yang akan digunakan dalam klasterisasi menggunakan pada data jumlah kejahatan menurut jenis kejahatan Provinsi di Indonesia tahun 2023 dengan menggunakan metode Kohonen SOMs yaitu pada ukuran *cluster* 3, yang mana hasil validasi internalnya memiliki akurasi terbaik.

IV. KESIMPULAN

Penelitian ini menggunakan metode *Self Organizing Maps* (SOMs), kita dapat mengelompokkan provinsi-provinsi di Indonesia berdasarkan jenis dan jumlah kejahatan yang terjadi. Analisis ini menunjukkan bahwa provinsi-provinsi tersebut dapat dibagi menjadi 3 kelompok besar. Pembagian ini didasarkan pada hasil evaluasi menggunakan beberapa metode, seperti *Dunn*, *Connectivity*, dan *Silhouette*. Hasil *cluster* yang terbentuk pada pengelompokkan provinsi di Indonesia berdasarkan indikator jumlah kejahatan menurut jenis kejahatan sebanyak 3 *cluster*. *Cluster* 1 tingkat jumlah

DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

kejahatan berdampak rendah terdiri dari 29 provinsi. *Cluster* 2 dengan dampak sedang terdiri dari 3 provinsi dengan menunjukkan tingkat kasus kejahatan paling dominan yaitu pada kejahatan terkait penipuan, penggelapan, penyeludupan & korupsi dibandingkan *cluster* lainnya. *Cluster* 3 dengan dampak tinggi terdiri dari 2 provinsi dengan karakteristik yang paling menonjol dengan menunjukkan hampir semua indikator jumlah kejahatan menurut jenis kejahatan mengalami rataan kasus kejahatan yang tertinggi daripada *cluster* lainnya.

DAFTAR PUSTAKA

Alamtaha, Z., Djakaria, I., & Yahya, N. I. 2023. Implementasi Algoritma Hierarchical Clustering dan Non-Hierarchical Clustering untuk Pengelompokkan Pengguna Media Sosial. *ESTIMASI: Journal of Statistics and Its Application*, 33-43.

Brock, G., Pihur, V., Datta, S., & Datta, S. 2008. clValid: Paket R untuk validasi cluster. *Jurnal Perangkat Lunak Statistik*, *25*, 1-22.

Dewi, D. A. I. C., & Pramita, D. A. K. 2019. Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi dan Informatika*, *9*(3), 102-109.

Fawaz, F., Fitriasari, N. S., & Rosalia, A. A. 2022. Perbandingan Algoritma Self Organizing Map dan Fuzzy C-Means dalam clustering hasil produksi ikan PPN Karangantu. *Explore: Jurnal Sistem Informasi dan Telematika (Telekomunikasi, Multimedia dan Informatika)*, *13*(2), 102-109.

Han, Jiawei, Jian Pei, and Hanghang Tong. 2022. *Data Mining: Concepts and Techniques*. Morgan kaufmann.

Harun, Haerul. 2015. "Pengelompokan Data DIPA Berbasis Penyerapan Anggaran Menggunakan Metode Self Organizing Map (SOM)." : 226–32.

Khormarudin, Agus Nur. 2016. "Teknik Data Mining: Algoritma K-Means Clustering." *J. Ilmu Komput*: 1–12.

Kohonen, Teuvo. 1989. *Self-Organizing Feature Maps*. Springer.

Muhidin, Asep. 2017. "Analisa Metode Hierarchical Clustering Dan K-Mean Dengan Model Lrfmp Pada Segmentasi Pelanggan." *Jurnal SIGMA* 8(3): 237–44.

Munawar, Ghifari, and K Kunci. 2015. "Implementasi Algoritma Self Organizing Map (SOM) Untuk Clustering Mahasiswa Pada Matakuliah Proyek (Studi Kasus: JTK POLBAN)." In *Prosiding Industrial Research Workshop and National Seminar*, , 66–78.

Shieh, Shu-Ling, and I-En Liao. 2012. "A New Approach for Data Clustering and Visualization Using Self-Organizing Maps." *Expert Systems with Applications* 39(15): 11924–33.

Simamora, B. 2005. *Analisis multivariat pemasaran*. Gramedia Pustaka Utama.

Simatupang, Nursariani, and Faisal. 2017. CV. Pustaka Prima *Kriminologi: Suatu Pengantar*. http://repository.umsu.ac.id/handle/123456789/15406.

Talakua, Mozart W, Zeth A Leleury, and A W Taluta. 2017. "Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014." *BAREKENG: Jurnal Ilmu Matematika dan Terapan* 11(2): 119–28.

Wicaksono, A. E. 2017. Implementasi Data Mining Dalam Pengelompokan Data Peserta Didik di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus Sman 16 Bekasi). *Jurnal Ilmiah Teknologi dan Rekayasa*, *21*(3).

DOI: https://doi.org/10.24036/ujsds/vol3-iss1/334