KLASTERISASI KRIMINALITAS DI INDONESIA MENGGUNAKAN METODE KOHONEN-SELF ORGANIZING MAP DAN PARTICLE OPTIMIZATION SWARM

Wenny Maria Tampubolon, Amri Muhaimin, Trimono
Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jl. Raya Rungkut Madya, Gunung Anyar, Gununganyar, Surabaya, Jawa Timur
amri.muhaimin.stat@upnjatim.ac.id

ABSTRAK

Kriminalitas merupakan perilaku yang melanggar hukum serta menyimpang dari norma sosial yang berlaku, dan dapat menimbulkan kerugian baik secara fisik, ekonomi maupun psikologis. Data BPS tahun 2024 mencatat peningkatan jumlah kasus kriminal dari 327.965 pada 2022 menjadi 584.991 pada 2023, sementara itu kasus yang berhasil diselesaikan hanya 299.517 kasus. Selisih yang cukup besar tersebut menunjukkan masih adanya kelemahan dalam penanganan kejahatan, sehingga diperlukan analisis pola kriminalitas untuk memperoleh gambaran kondisi antar provinsi di Indonesia. Penelitian ini bertujuan mengelompokkan provinsi di Indonesia berdasarkan data kriminalitas menggunakan metode *Kohonen Self-Organizing Map (SOM)*. Metode SOM mampu mengelompokkan data tanpa pengawasan *(unsupervised learning)* ke dalam representasi dua dimensi sehingga pola antarprovinsi dapat terlihat dengan lebih jelas. Untuk meningkatkan kinerja *SOM*, digunakan *Particle Swarm Optimization (PSO)* sebagai metode optimasi yang meniru perilaku sosial sekumpulan partikel untuk menemukan kombinasi parameter terbaik. Hasil optimasi dengan PSO menunjukkan parameter optimal *SOM* adalah grid 2×1, sigma 0,3, dan learning rate 0,9. Dengan parameter tersebut, terbentuk dua klaster dengan nilai *Silhouette Coefficient* sebesar 0,59 yang menandakan kualitas klasterisasi cukup baik. Temuan ini diharapkan dapat memperlihatkan perbedaan karakteristik kriminalitas antarprovinsi dan menjadi dasar penyusunan strategi penanganan yang lebih tepat sasaran.

Kata kunci : *Data Kriminalitas; Clustering; PSO; Kohonen-SOM; Silhouette Coefficient*

1. PENDAHULUAN

Kriminalitas masih menjadi salah satu masalah sosial yang cukup besar di Indonesia [1]. Secara umum, kriminalitas atau tindak kejahatan dapat dipahami sebagai berbagai bentuk perilaku yang melanggar hukum, aturan, maupun nilai yang berlaku dalam masyarakat. Berbagai bentuk pelanggaran hukum ini muncul karena banyak faktor pemicunya, mulai dari tekanan ekonomi, ketimpangan sosial, rendahnya tingkat pendidikan, hingga lingkungan yang kurang mendukung [2].

Menurut data dari Bisnis Indonesia Resources Center (BIRC) tahun 2023, Indonesia menempati peringkat kedua dalam indeks kejahatan di Asia Tenggara dengan skor 6,85 poin. Catatan Statistik Kriminal BPS tahun 2024 juga menunjukkan adanya peningkatan kasus yang cukup besar dari 327.965 kasus pada tahun 2022 menjadi 584.991 kasus pada tahun 2023. Sementara kasus yang terselesaikan baru mencapai sekitar 299.517 kasus. Selisih besar antara jumlah kasus yang terjadi dan yang terselesaikan ini menunjukkan masih adanya kelemahan dalam penganganan kejahatan di Indonesia.

Tingginya kasus kejahatan dan rendahnya jumlah kasus yang terselesaikan menunjukkan bahwa persoalan kriminalitas tidak hanya terkait dengan frekuensi kejahatan, tetapi juga dengan belum optimalnya pemahaman mengenai pola kejahatan terbentuk antar provinsi. Setiap provinsi memiliki ciri kriminalitas yang berbeda, namun perbedaan tersebut belum sepenuhnya tergambar dalam bentuk kelompok yang jelas. Tanpa gambaran pola yang lebih rinci, upaya penanganan sering kali berjalan tanpa melihat karakteristik khusus tiap wilayah.

Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik kriminalitasnya. Dengan klasterisasi tersebut, diharapkan muncul gambaran yang lebih terstruktur mengenai pola kriminalitas antarwilayah sehingga dapat menjadi dasar bagi penyusunan kebijakan penanggulangan kejahatan yang lebih sesuai dengan kebutuhan tiap daerah. Untuk mencapai tujuan tersebut, penelitian memanfaatkan pendekatan analisis berbasis data dengan teknik data mining.

Data mining mampu menggali informasi tersembunyi dari kumpulan data yang besar untuk menemukan pola atau kecenderungan tertentu. Dalam penelitian ini, teknik yang digunakan adalah metode *clustering* atau pengelompokan data, yang berfungsi untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan kesamaan karakteristik kriminalitasnya [3]. Metode yang digunakan adalah Self-Organizing Map (SOM).

Self-Organizing Map (SOM) merupakan salah satu metode *unsupervised learning* yang efektif dalam melakukan klasterisasi data. SOM bekerja dengan memetakan data berdimensi tinggi ke dalam bentuk representasi yang lebih sederhana tanpa menghilangkan informasi penting [4]. Namun, hasil dari SOM sangat dipengaruhi oleh parameter awal, seperti ukuran *grid*, nilai *learning rate*, dan *sigma*. Jika parameternya kurang tepat, hasil klasterisasinya bisa kurang optimal atau bahkan tidak mewakili data sebenarnya.

Untuk mengatasi hal tersebut, digunakan metode *Particle Swarm Optimization* (PSO). PSO merupakan algoritma optimasi yang terinspirasi dari perilaku kelompok, seperti kawanan burung atau ikan, dalam mencari posisi terbaik [5]. Dengan menerapkan PSO, parameter terbaik untuk SOM dapat ditemukan secara otomatis sehingga hasil klasterisasi menjadi lebih stabil dan akurat. Kombinasi kedua metode ini diharapkan dapat memberikan hasil pengelompokan yang lebih baik.

Berdasarkan latar belakang tersebut, peningkatan jumlah kasus kejahatan, ketidakseimbangan antara kasus yang terjadi dan yang terselesaikan, serta beragamnya karakter kriminalitas di tiap provinsi menunjukkan perlunya pendekatan yang lebih terarah dalam memahami pola kriminalitas di Indonesia. Dengan menggabungkan teknik klasterisasi SOM dan optimasi PSO, penelitian ini diharapkan mampu menghadirkan gambaran kriminalitas antardaerah yang lebih jelas serta mendukung penyusunan kebijakan yang lebih tepat sasaran.

2. TINJAUAN PUSTAKA

Klasterisasi adalah teknik untuk mengelompokkan data berdasarkan kemiripan karakteristik sehingga objek yang serupa berada dalam kelompok yang sama. Salah satu metode yang banyak dipakai adalah Self-Organizing Map (SOM), yang memetakan data berdimensi tinggi ke bentuk yang lebih sederhana agar pola di dalamnya lebih mudah dilihat. Kualitas hasil SOM dapat ditingkatkan dengan Particle Swarm Optimization (PSO), yang membantu menemukan parameter terbaik agar pemetaan lebih tepat. Berbagai penelitian sebelumnya juga menunjukkan bahwa metode seperti SOM efektif digunakan untuk menganalisis data sosial maupun kriminal.

Midyanti dan Bahri [6] menerapkan Self-Organizing Maps (SOM) untuk mengelompokkan kabupaten/kota di Kalimantan Barat berdasarkan Indeks Pembangunan Manusia (IPM) dengan menambahkan variabel kepadatan penduduk, jumlah guru dan murid, serta jumlah pengangguran. Metode ini menggunakan *learning rate* 0,001 dan maksimum iterasi 1100, menghasilkan empat cluster dengan nilai *Silhouette Coefficient* terbaik 0,331611. Hasil *profiling* menunjukkan karakteristik cluster yang berbeda pada setiap penambahan variabel. Keunggulan metode SOM dalam penelitian ini adalah kemampuannya memetakan data berdimensi tinggi menjadi dua dimensi yang mudah diinterpretasikan, namun hasilnya sensitif terhadap pemilihan parameter normalisasi dan jumlah variabel tambahan.

Imani et al. [7] menggunakan SOM untuk mengelompokkan kabupaten/kota di Provinsi Nusa Tenggara Timur berdasarkan sembilan indikator sosial, termasuk IPM, tingkat pengangguran, PDRB, dan angka buta huruf. Hasilnya menunjukkan empat cluster dengan variasi karakteristik sosial yang mencerminkan kondisi pascapandemi Covid-19. Penelitian ini menekankan bahwa metode SOM efektif untuk mengidentifikasi kesenjangan sosial di daerah dengan variabilitas tinggi, meskipun pemilihan variabel yang terlalu banyak dapat memengaruhi kejelasan interpretasi.

Alkhalidi et al. [8] menerapkan SOM dalam Sistem Informasi Geografis (SIG) untuk memetakan wilayah penyalahgunaan narkoba di Kabupaten Aceh Tenggara. Proses *clustering* membentuk empat cluster dengan karakteristik dan warna berbeda untuk tiap kelompok, kemudian divisualisasikan menggunakan ArcGIS 10.1. Hasil penelitian memberikan gambaran spasial distribusi penyalahgunaan narkoba yang membantu BNK setempat dalam pengambilan keputusan. Kelebihan pendekatan ini adalah integrasi SOM dengan GIS yang memungkinkan visualisasi yang intuitif, tetapi hasilnya bergantung pada kualitas data dan pemilihan parameter *learning rate*.

Bekti et al. [9] mengembangkan sistem informasi berbasis web untuk pemetaan wilayah berdasarkan *clustering* tingkat kerentanan kriminalitas di Daerah Istimewa Yogyakarta, Jawa Tengah, dan Jawa Timur menggunakan kombinasi metode SOM dan K-Means. Sistem dibangun dengan CMS WordPress dan pemetaan dilakukan melalui QGIS, menghasilkan visualisasi peta kerawanan kriminalitas yang interaktif. Kelebihan metode ini terletak pada integrasi data spasial dengan pendekatan statistik modern yang memudahkan interpretasi bagi pengguna umum, namun implementasinya membutuhkan proses pemeliharaan sistem yang cukup kompleks.

Inayah et al. [10] menerapkan algoritma Fuzzy C-Means (FCM) untuk mengelompokkan daerah rawan kriminalitas di Indonesia berdasarkan data BPS tahun 2018–2021. Hasilnya menghasilkan tiga cluster (tinggi, sedang, rendah) dengan nilai *Silhouette Coefficient* rata-rata 0,8322, menunjukkan kualitas klaster yang sangat baik. Penelitian ini membuktikan bahwa FCM mampu mengakomodasi ketidakpastian data dan memberikan hasil yang stabil. Kelebihannya adalah fleksibilitas dalam menangani data dengan derajat keanggotaan ganda, namun kelemahannya terletak pada kebutuhan komputasi yang lebih tinggi dibanding metode *hard clustering*.

3. METODE PENELITIAN

Dalam penelitian ini terdapat lima tahapan yang dilakukan, yaitu pengumpulan data, persiapan data, peneraparan optimasi, pengelompokkan data, dan evaluasi model. Setiap tahapan disusun secara sistematis untuk memastikan data diolah secara akurat dan menghasilkan kelompok atau klaster yang sesuai dengan karakteristik data.

3.1. Pengumpulan Data

Tahap pertama dalam penelitian adalah mengumpulkan data. Data yang digunakan merupakan data sekunder tahun 2023 yang berkaitan dengan data kriminalitas nasional. Data tersebut diperoleh dari publikasi Badan Pusat Statistik (BPS) berjudul

Statistik Kriminal 2024. Indikator yang dikaji meliputi (X_1) jumlah kejahatan terhadap nyawa, (X_2) jumlah kejahatan terhadap fisik, (X_3) jumlah kejahatan terhadap kesusilaan, (X_4) jumlah kejahatan terhadap hak milik, (X_5) jumlah kejahatan terkait narkotika, (X_6) jumlah kejahatan terkait penipuan, (X_7) jumlah kejahatan yang berhasil diselesaikan, (X_8) selang waktu terjadinya kejahatan dalam hitungan detik, serta (X_9) risiko penduduk terkena kejahatan.

Provinsi	x1	x2	x3	x4	x5	x6	x7	x8	x9
Aceh	18	1651	320	2948	1476	1357	10708	2539	216
Sumatera Utara	229	10381	752	23427	5308	8007	33385	506	433
Sumetera Barat	30	2490	288	2670	1274	1149	8821	2478	235
Riau	58	2041	363	5653	1985	1886	2870	1998	237
Jambi	26	1275	98	1926	735	976	3317	643	192

Gambar 1. Data Kriminalitas

3.2. Persiapan Data

Tahap kedua dalam penelitian adalah persiapan data. Proses persiapan data, dikenal juga sebagai *preprocessing*, bertujuan untuk menyesuaikan dan membersihkan data agar siap digunakan dalam analisis dan pemodelan. Tahapan persiapan data ini mencakup dua langkah, yaitu:

a. Pembersihan Data (*Data Cleaning*)
   Data mentah akan diperiksa apakah terdapat missing value dan data duplikat**.** Jika terdapat missing value, data akan ditangani melalui proses imputasi agar tidak menimbulkan bias pada analisis selanjutnya. Kemudian, dilakukan juga pemeriksaan terhadap data yang terduplikasi, dan jika terdapat baris data yang sama, baris tersebut akan dihapus [11]. Tahap ini dilakukan untuk memastikan bahwa data yang digunakan bersih, konsisten, dan siap diolah pada tahap pemodelan [12].
b. Standarisasi data
   Proses standarisasi dilakukan untuk menyamakan skala setiap variabel dalam dataset agar dapat digunakan secara optimal dalam pemodelan. Standarisasi bertujuan untuk menghindari perbedaan rentang nilai yang terlalu besar antar variabel, karena hal tersebut dapat memengaruhi hasil analisis dan proses klasterisasi. Penelitian ini menggunakan *Z-Score normalization*, di mana setiap nilai pada dataset ditransformasi berdasarkan rata-rata dan standar deviasinya [13], berikut rumus dari *Z-Score normalization* ditunjukkan pada persamaan (1).

$$Z_{uv} = \frac{x_{uv} - \bar{x}_v}{s_v} \quad (1)$$

Keterangan,
Z_{uv} : *Z-score* untuk data ke-u pada variabel ke- v
X_{uv} : Nilai asli data ke-u dalam variabel ke- v
$\bar{x}_v$: Nilai rata-rata dalam variabel ke- v
s_v : Standar deviasi variabel ke- v

3.3. *Particle Swarm Optimization (PSO)*

Tahap ketiga dalam penelitian adalah penerapan metode optimasi. Metode yang digunakan adalah *Particle Swarm Optimization (PSO)*, sebuah algoritma optimasi yang terinspirasi dari perilaku sekelompok burung ketika mencari makan. Dalam konsep ini, setiap burung diibaratkan sebagai partikel yang menjelajahi ruang solusi. Setiap partikel bergerak sambil berinteraksi dan saling berbagi informasi mengenai posisi terbaik yang pernah mereka temukan. Melalui proses tersebut, PSO berupaya menemukan solusi terbaik secara global dengan memilih dan memperbarui atribut-atribut partikel pada setiap iterasi [14].

Algoritma Particle Swarm Optimization (PSO) pertama kali diperkenalkan oleh Dr. Eberhart dan Dr. Kennedy pada tahun 1995. PSO termasuk ke dalam algoritma kecerdasan buatan, khususnya dalam kategori swarm intelligence, karena terinspirasi dari perilaku kolektif makhluk hidup seperti burung dan ikan dalam bertahan hidup. Sejak dikembangkan, PSO telah mengalami banyak perkembangan, baik dari sisi aplikasi maupun modifikasi algoritma. Kepopuleran metode ini disebabkan oleh karakteristiknya yang mudah dipahami, implementasinya sederhana, dan performanya yang terbukti handal dalam berbagai permasalahan optimasi [5].

Tahapan awal pada metode Particle Swarm Optimization (PSO) dimulai dengan menginisialisasi setiap partikel secara acak dari populasi awal yang berisi nilai posisi dan kecepatan (velocity). Setelah itu, ditentukan jumlah iterasi maksimum serta nilai bobot inersia yang akan mengatur seberapa besar pengaruh kecepatan sebelumnya terhadap pergerakan partikel berikutnya.

Selanjutnya, setiap partikel dievaluasi menggunakan fungsi *fitness* untuk mengukur seberapa baik solusi yang dihasilkan. Jika nilai *fitness* suatu partikel lebih baik dari nilai terbaik yang pernah dicapai sebelumnya, maka nilai p_best atau *personal best* dari partikel tersebut diperbarui. Setelah itu, sistem akan menentukan partikel dengan nilai *fitness* tertinggi di antara seluruh populasi dan memperbarui nilai g_best atau *global best* sesuai hasil tersebut.

Perubahan kecepatan setiap partikel kemudian dihitung dengan mempertimbangkan tiga komponen utama, yaitu kecepatan sebelumnya, jarak terhadap posisi terbaik partikel itu sendiri (*p_best*), dan jarak terhadap posisi terbaik seluruh partikel (*g_best*). Perubahan kecepatan partikel dihitung menggunakan persamaan (2) berikut:

$$v_{i,j}(t+1) = \omega v_{i,j}(t) + c_1 rand(\,)\,(p_{i,j} - h_{i,j}) + c_2 rand(\,)\,(g_j - h_{i,j}) \quad (2)$$

Keterangan,
$v_{i,j}(t)$: *Velocity* partikel ke-i pada dimensi ke-j di iterasi ke-t
$h_{i,j}$: Posisi partikel ke-i pada dimensi ke-j saat ini

$p_{i,j}$: Posisi *personal best* partikel ke-i pada dimensi ke-j
g_j : Posisi *global best* pada dimensi ke-j
ω : *Inertia weight*, mengontrol pengaruh kecepatan sebelumnya
c_1 : *Cognitive coefficient*, pengaruh pengalaman individu partikel
c_2 : *Social coefficient*, pengaruh partikel terbaik di seluruh populasi
$rand()$: Bilangan acak (0–1) untuk memberi efek eksplorasi

Selanjutnya, posisi setiap partikel diperbarui berdasarkan nilai kecepatan terbarunya, seperti ditunjukkan pada persamaan (3) berikut:

$$p_{i,j}(t+1) = p_{i,j}(t) + v_{i,j}(t+1) \quad (3)$$

Keterangan,
$p_{i,j}(t)$: Posisi partikel ke-*i* pada dimensi ke-*j* di iterasi ke-*t*
$p_{i,j}(t+1)$: Posisi partikel ke-*i* pada dimensi ke-*j* di iterasi berikutnya (t+1)
$v_{i,j}(t+1)$: Kecepatan partikel ke-*i* pada dimensi ke-*j* setelah diperbarui

Setelah posisi diperbarui, nilai p_{best} juga diperbarui menggunakan persamaan (4) berikut:

$$p_{best}(t+1) = \begin{cases} p_{best}(t), \ jika\ f(h(t+1)) \le f(p_{best}(t)) \\ h(t+1), \ jika\ f(h(t+1)) > f(p_{best}(t)) \end{cases} \quad (4)$$

Keterangan,
$p_{best}(t)$: Posisi terbaik partikel hingga iterasi ke-t (*personal best*)
$p_{best}(t+1)$: Posisi *personal best* yang diperbarui pada iterasi berikutnya
$h(t+1)$: Posisi partikel saat ini pada iterasi ke-(t+1)
$f(.)$: Fungsi *fitness* yang digunakan untuk menilai kualitas solusi

Setelah setiap partikel memperbarui posisinya, langkah berikutnya adalah menentukan nilai g_best atau *global best*. Nilai ini diperoleh dengan mencari partikel yang memiliki nilai *p_best* tertinggi di antara seluruh partikel dalam populasi, yang menandakan solusi terbaik secara keseluruhan pada iterasi tersebut. Proses pembaruan ini memastikan bahwa seluruh partikel terus bergerak menuju arah solusi global yang paling optimal.

Langkah-langkah perhitungan ini kemudian diulangi secara berurutan, mulai dari evaluasi nilai *fitness*, pembaruan *p_best* dan *g_best*, hingga perhitungan posisi dan kecepatan partikel, selama jumlah iterasi maksimum yang telah ditentukan belum tercapai. Siklus ini berlangsung terus-menerus sampai algoritma mencapai kondisi konvergensi atau tidak ada lagi peningkatan signifikan pada nilai *fitness*. Hasil akhirnya adalah posisi partikel dengan nilai *fitness* tertinggi, yang dianggap sebagai solusi terbaik dari proses optimasi menggunakan metode Particle Swarm Optimization (PSO)

3.4. *Self Organizing Map (SOM)*

Tahap kelima dalam penelitian adalah pengelompokkan data atau *clustering*. Pada tahap ini digunakan algoritma *Self Organizing Map (SOM)*, yaitu metode dalam kecerdasan buatan yang memetakan data berdimensi tinggi ke dalam grid atau peta topologi [4]. SOM berfungsi menyederhanakan representasi data tanpa menghilangkan informasi penting. Teknik ini mengelompokkan data berdasarkan pola dan hubungan antar variabel, sehingga efektif digunakan dalam analisis *cluster* maupun visualisasi data.

Algoritma SOM pertama kali dikembangkan pada tahun 1980-an oleh Teuvo Kohonen, seorang ahli di bidang kecerdasan buatan [15]. Metode ini termasuk dalam kategori *Artificial Neural Network (ANN)* dengan pendekatan *unsupervised learning*, sehingga mampu belajar secara mandiri tanpa memerlukan label atau keluaran yang telah ditentukan sebelumnya [15]. Salah satu keunggulan utama SOM adalah kemampuan dalam mereduksi dimensi data sambil mempertahankan struktur hubungan antar variabel.

Adapun proses SOM meliputi beberapa tahapan, tahapan awal dalam metode *Self-Organizing Map* (SOM) dimulai dengan menentukan sejumlah parameter yang dibutuhkan sebelum proses pelatihan dimulai. Parameter tersebut meliputi bobot awal w_{uv}, radius tetangga (R), *learning rate*, koefisien penurun *learning rate*, serta jumlah *epoch* atau iterasi pelatihan yang akan dijalankan. Setelah parameter ditentukan, proses pelatihan dilakukan secara berulang hingga jumlah iterasi yang telah ditetapkan tercapai.

Pada setiap iterasi, setiap data masukan x_u akan diproses secara berurutan melalui beberapa tahapan. Pertama, dilakukan perhitungan jarak antara data masukan dengan bobot masing-masing neuron pada jaringan. Perhitungan jarak ini bertujuan untuk menemukan neuron yang memiliki bobot paling mendekati data masukan, atau yang disebut *Best Matching Unit* (BMU). Jarak tersebut dihitung menggunakan persamaan (5) berikut:

$$d_v = \sum_{u=1}^{n} (w_{uv} - x_u)^2 \quad (5)$$

Keterangan,
d_v : jarak antara neuron dengan data masukan
n : jumlah iterasi
u : indeks iterasi awal
w_{uv} : bobot yang digunakan dalam perhitungan
x_u : nilai data masukan

Setelah jarak antara setiap data masukan dan neuron dihitung, langkah berikutnya adalah menentukan neuron dengan jarak terkecil. Neuron ini disebut *Best Matching Unit* (BMU), yaitu neuron yang

memiliki bobot paling mendekati data masukan dan dianggap paling mewakili pola data tersebut.

Selanjutnya, dilakukan proses pembaruan bobot pada neuron BMU dan neuron-neuron lain yang berada dalam radius tetangganya. Pembaruan ini bertujuan agar bobot neuron-neuron tersebut semakin mendekati nilai data masukan. Prosesnya dilakukan berdasarkan nilai radius yang telah ditentukan sebelumnya, dengan mempertimbangkan pengaruh *learning rate* yang akan menurun secara bertahap seiring bertambahnya iterasi. Pembaruan bobot dapat dihitung dengan persamaan (6) berikut:

$$w_{uv}(new) = w_{uv}(old) + \alpha[x_u - w_{uv}(old)] \quad (6)$$

Keterangan,
$w_{uv}(new)$: bobot yang diperbarui setelah proses pembelajaran
$w_{uv}(old)$: bobot sebelum diperbarui
α : *learning rate*
x_u : data masukan yang digunakan dalam perhitungan

Setelah bobot neuron diperbarui, langkah selanjutnya adalah menyesuaikan nilai *learning rate*. Nilai ini menggambarkan seberapa besar perubahan bobot yang dilakukan pada setiap iterasi. Pada tahap awal pelatihan, nilai learning rate biasanya cukup besar agar jaringan dapat beradaptasi dengan cepat terhadap pola data. Namun, seiring berjalannya proses, nilai tersebut akan diperbarui menggunakan rumus pada persamaan (7) dan menurun secara bertahap.

$$\alpha(baru) = \beta * \alpha(lama) \quad (7)$$

Keterangan,
$\alpha(baru)$: *learning rate* baru
$\alpha(lama)$: *learning rate* sebelumnya
β : koefisien penurun *learning rate*

3.5. Evaluasi Model

Tahap terakhir dalam penelitian adalah evaluasi model. Setelah hasil dari penerapan metode *SOM* dan optimasi PSO selesai, dilakukan evaluasi hasil clustering untuk memastikan bahwa data telah dikelompokkan dengan baik. Pada penelitian ini menggunakan metode evaluasi yang umum digunakan, yaitu *Silhouette Coefficient*. *Silhouette Coefficient* mengukur seberapa dekat suatu data dengan *clusternya* sendiri dibandingkan dengan *cluster* lain. Nilai *Silhouette* berkisar antara -1 hingga 1, nilai yang mendekati 1 menunjukkan *clustering* yang lebih baik [13].

Tahap awal dalam perhitungan Silhouette Coefficient adalah menghitung rata-rata jarak antara objek u dengan semua objek lain dalam *cluster* yang sama, seperti ditunjukkan pada persamaan (8) berikut:

$$a(u) = \frac{1}{n_A - 1} \sum_{v \in A, v \neq u} d(uv) \quad (8)$$

Keterangan,
$a(u)$: Rata-rata jarak objek u ke objek lain dalam *cluster* A
n_A : Total jumlah objek yang terdapat dalam *cluster* A
$d(uv)$: Jarak antara objek u dan objek v
$v \in A$: Objek lain dalam *cluster* A yang berbeda dari u
$v \neq u$: Objek v berbeda dari objek u

Selanjutnya menghitung rata-rata jarak antara objek ke- u dan seluruh objek di *cluster* lain, lalu pilih nilai yang paling kecil, seperti yang ditunjukkan persamaan (9).

$$d(u, C) = \frac{1}{n_C} \sum_{v \in C} d_{uv} \quad (9)$$

Keterangan,
n_C : Jumlah anggota dalam *cluster* C
$d(u, C)$: Rata-rata jarak objek ke-u dengan semua objek *cluster* lain
d_{uv} : Jarak antara objek ke-u dan objek ke-v
$v \in A$: Objek ke-v termasuk dalam *cluster* C

Setelah itu memilih nilai terkecil dari semua kelompok dengan C ≠ A, seperti pada persamaan (10).

$$b(u) = min_{u \in C} d(u, C) \quad (10)$$

Keterangan,
$b(u)$: jarak minimal dari $d(u, C)$
$u \in C$: Objek u termasuk dalam *cluster* C
$d(u, C)$: Rata-rata jarak objek u ke semua objek dalam *cluster* C

Tahap terakhir menghitung nilai *Silhouette Coefficientnya*, seperti yang ditunjukkan pada persamaan (11).

$$s(u) = \frac{b(u) - a(u)}{\max(a(u) - b(u))} \quad (11)$$

Keterangan,
$s(u)$: Koefisien *Silhouette* untuk data u
$b(u)$: Jarak terdekat dari objek u ke *cluster* lain.
$a(u)$: Rata-rata jarak objek u dengan semua objek dalam *cluster* yang sama

4. HASIL DAN PEMBAHASAN

4.1. Pembersihan Data (*Data Cleaning*)

Pada tahap ini, dilakukan pengecekan terhadap *missing value* dan data duplikat agar data yang digunakan bersih dan siap untuk dianalisis. Tabel 1 dan Tabel 2 menampilkan hasil dari *missing value* dan data duplikat.

Tabel 1. Missing Value

Variabel	Jumlah Nilai Hilang
X_1	0
X_2	0
X_3	0
X_4	0
X_5	0
X_6	0
X_7	0
X_8	0
X_9	0

Pada Tabel 1 setiap variabel menampilkan angka 0, hal ini menunjukkan bahwa tidak terdapat nilai yang hilang pada setiap variabel. Selanjutnya, dilakukan pengecekan terhadap data duplikat.

Tabel 2. Data Duplikat

Variabel	Jumlah Data Berduplikat
X_1	0
X_2	0
X_3	0
X_4	0
X_5	0
X_6	0
X_7	0
X_8	0
X_9	0

Pada Tabel 3 setiap variabel menampilkan angka 0, hal ini menunjukkan bahwa tidak terdapat data berduplikat pada setiap variabel.

4.2. Standarisasi Data

Pada tahap ini dilakukan standarisasi data, yakni sebuah proses mengubah skala nilai pada setiap variabel agar memiliki rata-rata (mean) 0 dan standar deviasi 1. Tujuannya agar setiap variabel memiliki skala yang sebanding, sehingga tidak ada variabel dengan nilai besar yang mendominasi proses analisis. Tabel 3 contoh data sebelum di standarisasi dan sesudah di standarisasi.

Table 3. Hasil Standarisasi Data

Provinsi	X_1	
	Data Asli	*Z-score*
Aceh	18	-0.633
Sumatera Utara	229	2.631
Sumetera Barat	30	-0.447
Riau	58	-.0.014
Jambi	26	-0.509
Sumatera Selatan	87	0.434

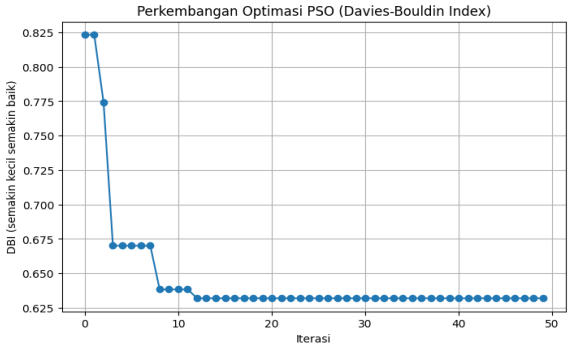
4.3. Optimasi Parameter dengan *Particle Swarm Optimization (PSO)*

Tahap optimasi dilakukan menggunakan *Particle Swarm Optimization (PSO)* untuk memperoleh kombinasi parameter terbaik pada model *Self-Organizing Map (SOM)*. PSO bekerja dengan membentuk sejumlah partikel yang masing-masing mewakili kandidat parameter berupa ukuran grid, nilai sigma, dan learning rate. Tabel 4 menunjukkan parameter optimal yang dihasilkan oleh PSO.

Table 4. Parameter Terbaik

Parameter	Nilai
Grid X x Y	2 x 1
Sigma	0,3
Learning Rate	0,9

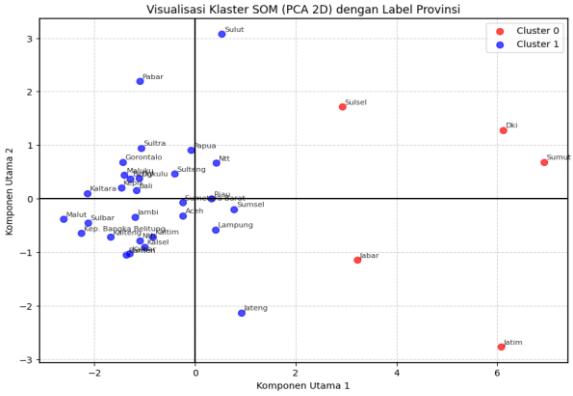
Parameter terbaik yang ditemukan PSO, yaitu *grid* 2 × 1, *sigma* 0.3, dan *learning rate* 0.9 menghasilkan nilai *Davies Bouldin Index* (DBI) terendah. Grafik pada gambar 2 menunjukkan nilai DBI turun pada iterasi awal, kemudian melandai dan stabil di sekitar 0.63. Pola ini menggambarkan bahwa PSO bergerak cepat menemukan kombinasi parameter yang lebih baik lalu berhenti berubah ketika mencapai titik optimal. DBI yang stabil di bagian akhir menunjukkan proses optimasi selesai dengan baik dan parameter tersebut digunakan sebagai parameter final pada pelatihan SOM.



Gambar 2. Nilai DBI

4.4. Pemodelan Kohonen-Self Organizing Map dan Evaluasi

Model SOM kemudian dilatih menggunakan parameter terbaik hasil optimasi PSO. Proses pelatihan dilakukan dengan menyesuaikan bobot neuron terhadap data hingga menemukan *Best Matching Unit* (BMU) untuk setiap provinsi. Hasil pengelompokkan tersebut membentuk 2 *cluster* dengan nilai *Silhouette Coefficient* sebesar 0,59, nilai evaluasi tersebut menunjukkan klaster dapat dikategorikan baik. Visualisasi klasterisasi ditunjukkan pada Gambar 3.

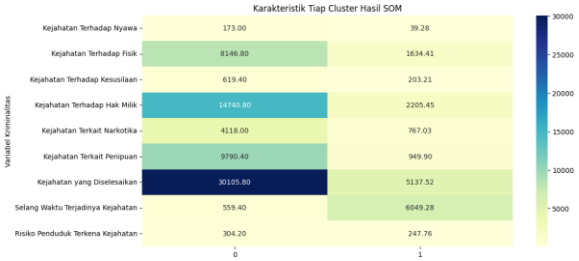


Gambar 3. Visualisasi Hasil Klaster

Pada Gambar 3 menunjukkan hasil pengelompokan provinsi menggunakan SOM yang kemudian dipetakan ke tampilan dua dimensi menggunakan PCA. Sumbu X (Komponen Utama 1) menunjukkan besar-kecilnya nilai variabel secara keseluruhan, sehingga provinsi yang berada semakin ke kanan memiliki nilai indikator yang lebih tinggi. Sementara itu, sumbu Y (Komponen Utama 2) menggambarkan perbedaan pola antarprovinsi; titik yang berada lebih tinggi menandakan adanya karakter

atau kecenderungan tertentu yang lebih menonjol dibanding provinsi lainnya.

Setiap titik pada grafik merepresentasikan satu provinsi, dan perbedaan warna menunjukkan klaster yang terbentuk. Klaster 0 yang ditandai dengan warna merah tampak mengelompok di bagian kanan grafik, menggambarkan provinsi dengan nilai indikator relatif tinggi. Sebaliknya, klaster 1 yang berwarna biru tersebar di area tengah hingga kiri, menandakan provinsi dengan tingkat indikator yang lebih rendah atau sedang. Provinsi-provinsi yang posisinya berdekatan menunjukkan pola data yang mirip, sedangkan provinsi yang saling berjauhan memiliki karakteristik yang berbeda lebih jelas. Setelah mengetahui jumlah klaster, Gambar 4 menjelaskan karakteristik tiap klaster.



Gambar 4. Karakteristik Tiap Cluster

Pada Gambar 5 menampilkan rata-rata nilai tiap variabel kriminalitas untuk setiap klaster hasil SOM. Perbedaan warna membantu menunjukkan variabel mana yang memiliki nilai paling tinggi pada masing-masing kelompok. Klaster 0 memiliki nilai yang jauh lebih tinggi pada hampir semua jenis kejahatan, terutama kejahatan terhadap hak milik, penipuan, serta jumlah kasus yang berhasil diselesaikan. Hal ini menunjukkan bahwa provinsi dalam klaster ini memiliki tingkat kriminalitas yang tinggi sekaligus aktivitas penanganan kasus yang lebih besar.

Sedangkan klaster 1 menunjukkan nilai yang lebih rendah pada hampir seluruh variabel kriminalitas, menandakan bahwa provinsi dalam kelompok ini memiliki tingkat kejahatan yang rendah. Perbedaan warna pada visualisasi mempermudah melihat karakteristik utama kedua klaster. Melalui pola ini, penelitian dapat memahami perbedaan antar klaster dan variabel yang paling menentukan pembentukannya. Untuk melihat anggota tiap klaster secara lebih rinci, daftar provinsi pada masing-masing klaster ditampilkan pada Tabel 5 berikut.

Pada Tabel 5 klasterisasi menggunakan SOM menghasilkan dua klaster utama. Klaster 0, terdiri dari 5 provinsi berisi Sumatera Utara, DKI Jakarta, Jawa Barat, Jawa Timur, dan Sulawesi Selatan. Klaster ini memiliki nilai kriminalitas yang lebih tinggi dibanding provinsi lainnya, sehingga membentuk kelompok tersendiri yang memiliki karakteristik yang lebih menonjol dalam berbagai variabel kejahatan.. Variabel seperti kejahatan terhadap fisik, hak milik, penipuan, narkotika, serta jumlah kejahatan yang diselesaikan menunjukkan angka paling besar. Hal ini menjelaskan mengapa provinsi-provinsi dengan aktivitas kriminal yang padat dan kompleks ini terkumpul dalam satu klaster.

Table 5. Daftar Provinsi Tiap Cluster

Cluster	Nama Provinsi	Jumlah
0	Sumatera Utara, DKI Jakarta, Jawa Barat, Jawa Timur, dan Sulawesi Selatan.	5
1	Aceh, Sumatera Barat, Riau, Kep. Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Banten, Jawa Tengah, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Gorontalo, Sulawesi Tengah Sulawesi Barat, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat, dan Papua	29

Pada klaster 1, terdiri dari 29 provinsi yang mencakup Aceh, Sumatera Barat, Riau, dan lain-lain, memiliki nilai kriminalitas rendah. Hampir semua variabel kejahatan berada pada nilai minimal, dan selang waktu terjadinya kejahatan relatif lebih panjang, menandakan frekuensi kejahatan yang rendah. Karena itulah provinsi-provinsi dengan kondisi kriminalitas yang lebih ringan terkumpul dalam klaster ini

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa penerapan Self Organizing Map (SOM) yang dioptimasi menggunakan Particle Swarm Optimization (PSO) mampu mengelompokkan provinsi di Indonesia menjadi 2 klaster berdasarkan pola kriminalitasnya. Hasil klasterisasi ini memberikan gambaran yang lebih jelas mengenai wilayah dengan tingkat kriminalitas tinggi dan rendah sehingga dapat menjelaskan kesenjangan atau selisih antara jumlah kasus yang terjadi dan kasus yang berhasil diselesaikan seperti yang dijelaskan pada latar belakang. Informasi tersebut berpotensi mendukung pemerintah dan aparat penegak hukum dalam memprioritaskan wilayah intervensi serta merumuskan strategi penanganan yang lebih terarah, sehingga upaya pengurangan disparitas penyelesaian kasus dapat dilakukan secara lebih efektif. Penelitian selanjutnya disarankan untuk mencoba metode optimasi lainnya, seperti Genetic Algorithm, Bayesian Optimization, atau pendekatan hybrid, guna memperoleh parameter SOM yang lebih stabil dan akurat. Selain itu, lebih mudah untuk memahami pola kriminalitas dengan menambahkan metode pembelajaran lain atau memperluas variabel analisis. Sehingga hasil klasterisasi dapat lebih relevan untuk mendukung pengambilan kebijakan berbasis data.

DAFTAR PUSTAKA

[1] S. Miftahurrahmi, N. Zilrahmi, N. Amalita, and T. O. Mukhti, "Metode Density Based Spatial Clustering of Applications with Noise (DBSCAN) dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Kriminalitas Tahun 2022," UNP Journal of Statistics and Data Science, vol. 2, no. 3, pp. 330–337, Aug. 2024, doi: https://doi.org/10.24036/ujsds/vol2-iss3/203.

[2] N. Azizah, A. Fauzi, T. Rohana, and S. Faisal, "Perbandingan Metode K-Means dan K-Medoids Untuk Clustering Jenis Kriminalitas," Building of Informatics, Technology and Science (BITS), vol. 6, no. 2, pp. 1011–1019, 2024, doi: https://doi.org/10.47065/bits.v6i2.5723.

[3] N. Dicky, N. Y. Maulita, and N. S. Ramadani, "Pengelompokan Data Kriminal untuk Menentukan Pola Rawan Tindak Kriminal Menggunakan Algoritma K-Means," vol. 2, no. 5, pp. 122–138, Sep. 2024, doi: https://doi.org/10.62383/polygon.v2i5.238.

[4] A. Muhaimin and K. Fithriasari, 'Kohonen-SOM LOF Approach for Anomaly Detection," pp.1-6, Oct. 2021, doi: https://doi.org/10.1109/itis53497.2021.9791596.

[5] N. A. Febiani, N. A. M. Widodo, N. N. Anwar, N. B. A. Sekti, and N. A. Yulfitri, "Implementasi Algoritma 'Particle Swarm Optimization' (PSO) Penjadwalan Belajar Mengajar," IIKRA-ITH Informatika Jurnal Komputer dan Informatika, vol. 8, no. 1, pp. 152–161, Nov. 2023, doi: https://doi.org/10.37817/ikraith-informatika.v8i1.3210.

[6] D.M.Midyanti and S. Bahri, "Implementasi Self Organizing Maps untuk Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia," Jurnal Teknologi Informasi dan Ilmu Komputer, vol.10, no.6, pp. 1265-1272, Dec. 2023, doi: https://doi.org/10.25126/jtiik.2023107647.

[7] N. Imani, A. I. Alfassa, and A. M. Yolanda, "ANALISIS CLUSTER TERHADAP INDIKATOR DATA SOSIAL DI PROVINSI NUSA TENGGARA TIMUR MENGGUNAKAN METODE SELF ORGANIZING MAP (SOM)," Jurnal Gaussian, vol. 11, no. 3, pp. 458-467, Jan. 2023. https://doi.org/10.14710/j.gauss.11.3.458-467.

[8] M. Alkhalidi, B. Nadeak, and M. Sayuthi, "Sistem Informasi Geografis Pemetaan Wilayah Penyalahgunaan Narkoba Mengunakan Metode SOM (Self-Organizing Map) Studi Kasus: Kabupaten Aceh Tenggara", bits, vol. 2, no. 1, pp. 1-9, Jun. 2020.

[9] R. D. Bekti, R. N. Zulfahmi, M. K. Daul, W. J. Pradnyaana, and E. Sutanta, "SISTEM INFORMASI BERBASIS WEBSITE UNTUK PEMETAAN WILAYAH BERDASARKAN CLUSTERING KERENTANAN KRIMINALITAS", JINTEKS, vol. 6, no. 3, pp. 620-626, Aug. 2024.

[10] P. A. Riyantoko, T. M. Fahrudin, K. M. Hindrayani, and M. Idhom, "Exploratory Data Analysis and Machine Learning Algorithms to Classifying Stroke Disease," IJCONSIST JOURNALS, vol. 2, no. 02, pp. 77–82, Jun. 2021, doi: https://doi.org/10.33005/ ijconsist.v2i02.49.

[11] T. M. Fahrudin, P. A. Riyantoko, K. M. Hindrayani, and I. G. S. Mas Diyasa, "Exploratory Data Analysis pada Kasus COVID-19 di Indonesia Menggunakan HiveQL dan Hadoop Environment", santika, vol. 1, pp. 115–123, Nov. 2020.

[12] T. M. Fahrudin, P. A. Riyantoko, K. M. Hindrayani, and M. H. P. Swari, "Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators using Agglomerative Hierarchical Clustering", Telematika, vol. 18, no. 2, pp. 194–210, Oct. 2021.

[13] D. A. Prasetya, A. P. Sari, M. Idhom, and A. Lisanthoni, "Optimizing Clustering Analysis to Identify High-Potential Markets for Indonesian Tuber Exports", ijeeemi, vol. 7, no. 1, pp. 113–122, Feb. 2025, doi: 10.35882/skzqbd57.

[14] A. Febiani, A. M. Widodo, N.n Anwar, B. A. Sekti, and A. Yulfitri, "Implementasi Algoritma 'Particle Swarm Optimization' (PSO) Penjadwalan Belajar Mengajar", *ikraith-informatika*, vol. 8, no. 1, pp. 152-161, Mar. 2024.

[15] I. IRWAN, A. Y. HASHARI, H. IHSAN, and A. ZAKI, "PENGGUNAAN SELF ORGANIZING MAP DALAM PENGELOMPOKAN TINGKAT KESEJAHTERAAN MASYARAKAT," *Jambura Journal of Probability and Statistics*, vol. 1, no. 2, pp. 57–68, Sep. 2020, doi: https://doi.org/10.34312/jjps.v1i2.7266.