

**PENGGUNAAN METODE GAUSSIAN KERNEL FUZZY C-MEANS
UNTUK PENGELOMPOKAN KABUPATEN/KOTA PROVINSI SULAWESI
SELATAN BERDASARKAN DATA KRIMINALITAS**



**AYU LESTARI
H051191046**



**PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**PENGGUNAAN METODE *GAUSSIAN KERNEL FUZZY C-MEANS*
UNTUK PENGELOMPOKAN KABUPATEN/KOTA PROVINSI SULAWESI
SELATAN BERDASARKAN DATA KRIMINALITAS**

AYU LESTARI

H051191046



**PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**PENGGUNAAN METODE GAUSSIAN KERNEL FUZZY C-MEANS
UNTUK PENGELOMPOKAN KABUPATEN/KOTA PROVINSI SULAWESI
SELATAN BERDASARKAN DATA KRIMINALITAS**

AYU LESTARI

H051191046

Skripsi

Diajukan sebagai salah satu syarat untuk mencapai gelar Sarjana Statistika

Program Studi Statistika

pada

**PROGRAM STUDI STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

SKRIPSI
PENGGUNAAN METODE GAUSSIAN KERNEL FUZZY C-MEANS
UNTUK PENGELOMPOKAN KABUPATEN/KOTA PROVINSI SULAWESI
SELATAN BERDASARKAN DATA KRIMINALITAS

AYU LESTARI

H051191046

Skripsi,

Telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada 14 Agustus 2024
dan dinyatakan telah memenuhi syarat kelulusan

pada

Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:

Pembimbing tugas akhir,

Anisa, S.Si., M.Si.

NIP.1973022719988022001



Diketahui:
Ketua Program Studi,

Dr. Anna Lemiyati, S.Si., M.Si.

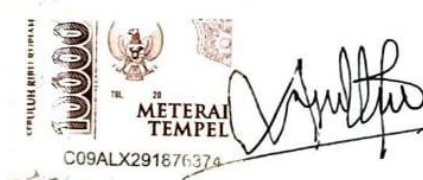
NIP.197708082005012002

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul “Penggunaan Metode *Gaussian Kernel Fuzzy C-Means* Untuk Pengelompokan Kabupaten/Kota Provinsi Sulawesi Selatan Berdasarkan Data Kriminalitas” adalah benar karya saya dengan arahan dari Ibu Anisa, S.Si., M.Si. sebagai Pembimbing. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 14 Agustus 2024



Ayu Lestari

NIM H051191046


UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Allah *Subhanahu Wa Ta'ala* atas segala limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam*, yang telah membawa kita dari zaman kegelapan menuju zaman yang terang benderang. *Alhamdulillahil'alamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanahu Wa Ta'ala*, penelitian ini dapat terlaksana dan terselesaikan atas bimbingan, diskusi dan arahan dari Ibu **Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.** selaku penasehat akademik dan ibu **Anisa, S.Si., M.Si.** selaku pembimbing yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya. Ucapan terima kasih juga penulis sampaikan kepada Ibu **Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.** dan ibu **Dra. Nasrah Sirajang, M.Si.** atas saran dan kritikan yang berharga serta waktu yang telah diberikan kepada penulis. Terima kasih juga kepada pimpinan Universitas Hasanuddin, Ketua Departemen Statistika, para dosen dan staff yang telah memberikan ilmu dan fasilitas kepada penulis.

Ucapan terima kasih serta penghargaan setinggi-tingginya penulis haturkan kepada keluarga terkhusus orang tua, Ayah **Zaenal Arifin** dan Ibu **Marwah** yang telah memberikan dukungan, pengorbanan kasih sayang serta doa yang tak henti-hentinya dipanjatkan kepada penulis.

Penulis juga berterima kasih kepada teman-teman terkhusus **Syamsul, Daus, Ferdi, Fadhel, Alya, Dian, Tamara, Wilda, Astri, Fitri, Fira, Nita**, yang telah memberikan semangat lebih kepada penulis dan cerita berharga dalam kehidupan bermahasiswa. Teman-teman **MIPA 2019**, yang menjadi tempat berbagi cerita, serta semua pihak yang tidak dapat disebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan bernilai ibadah disisi Allah *Subhanahu Wa Ta'ala*. Penulis menyadari bahwa masih banyak kekurangan, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat untuk pembaca.

Makassar, 14 Agustus 2024



AYU LESTARI

ABSTRAK

AYU LESTARI. **Penggunaan metode *gaussian kernel fuzzy c-means* untuk pengelompokan kabupaten/kota provinsi sulawesi selatan berdasarkan data kriminalitas** (dibimbing oleh Anisa).

Latar Belakang. Analisis *cluster* adalah analisis untuk mengelompokkan data yang memiliki elemen dan karakteristik serupa ke dalam kelompok (*cluster*) yang berbeda dengan mengklasifikasi berdasarkan variabel-variabel yang dipertimbangkan dalam penelitian. *Fuzzy C-Means* (FCM) yang merupakan salah satu metode *clustering non-hierarki* dengan keanggotaan data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. FCM menggunakan jarak antar data dalam menentukan nilai keanggotaan observasi, sehingga metode FCM sangat rentan terhadap pengaruh *noise* dan perbedaan skala nilai pada data. **Metode.** GKFCM bertujuan untuk meningkatkan kemampuan FCM dalam menangkap struktur dan perbedaan variabel yang lebih kompleks dalam data dengan memetakan data menggunakan *kernel*. **Tujuan.** Penelitian ini adalah untuk memperoleh jumlah *cluster* terbaik serta karakteristik setiap *cluster* pada data kriminalitas kabupaten/kota di Provinsi Sulawesi Selatan menggunakan GKFCM. **Hasil.** *Cluster* yang terbentuk sebanyak 4 kelompok wilayah berdasarkan nilai *Davies-Bouldin Index* terendah sebesar 0,82. Kelompok pertama merupakan wilayah dengan resiko kejahatan tertinggi diikuti oleh kelompok kedua dan keempat hingga kelompok ketiga yang memiliki resiko kejahatan paling rendah. Nilai evaluasi *Average Silhouette Value* metode GKFCM sebesar 0,74 menunjukkan bahwa kualitas *cluster* yang terbentuk termasuk ke dalam struktur *cluster* yang sangat baik. **Kesimpulan.** Metode GKFCM mampu mengelompokkan kabupaten/kota di Provinsi Sulawesi Selatan menjadi 4 *cluster* dengan kualitas yang sangat baik, yang menunjukkan perbedaan risiko kejahatan yang signifikan di antara kelompok-kelompok wilayah di Sulawesi Selatan.

Kata Kunci: *Average Silhouette Value, Fuzzy Clustering, Gaussian Kernel Fuzzy Clustering, Kriminalitas, Sulawesi Selatan*

ABSTRACT

AYU LESTARI. **The Use of Gaussian Kernel Fuzzy C-Means for Clustering Regencies/Cities in South Sulawesi Province Based on Crime Data** (supervised by Anisa).

Introduction. Cluster analysis is a method to group data with similar elements and characteristics into different clusters by classifying them based on variables considered in the research. Fuzzy C-Means (FCM) is one of the non-hierarchical clustering methods where data membership in a cluster is determined by the degree of membership. FCM uses the distance between data points to determine membership values, making it very susceptible to noise and differences in data value scales. **Method.** GKFCM aims to enhance FCM's ability to capture more complex structures and variable differences in data by mapping the data using a kernel. **Objective.** This research aims to obtain the best number of clusters and the characteristics of each cluster in crime data of regencies/cities in South Sulawesi Province using GKFCM. **Results.** The formed clusters consist of 4 groups of regions based on the lowest Davies-Bouldin Index value of 0.82. The first group represents areas with the highest crime risk, followed by the second and fourth groups, with the third group having the lowest crime risk. The evaluation value of the GKFCM method's Average Silhouette Value of 0.74 indicates that the formed clusters are of very good quality. **Conclusion.** The GKFCM method effectively clusters regencies/cities in South Sulawesi Province into 4 clusters with very good quality, demonstrating significant differences in crime risk among the region groups in South Sulawesi.

Keywords: Average Silhouette Value, Crime, Fuzzy Clustering, Gaussian Kernel Fuzzy Clustering, South Sulawesi

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN PENGAJUAN	ii
HALAMAN PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI	iv
UCAPAN TERIMA KASIH	v
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR NOTASI	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Landasan Teori	3
1.6.1 Analisis <i>Cluster</i>	3
1.6.2 <i>Fuzzy C-Means Clustering</i>	3
1.6.3 <i>Gaussian Kernel</i>	6
1.6.4 <i>Gaussian Kernel Fuzzy C-Means Clustering</i>	6
1.6.5 <i>Davies-Bouldin Index</i>	7
1.6.6 <i>Average Silhouette Value</i>	8
1.6.7 Kriminalitas.....	10
BAB II METODOLOGI PENELITIAN	12
2.1 Jenis dan Sumber Data	12
2.2 Struktur Data.....	12
2.3 Metode Analisis.....	12
BAB III HASIL DAN PEMBAHASAN.....	14

3.1 Penentuan Jumlah <i>Cluster</i> Menggunakan <i>Davies-Bouldin Index</i>	14
3.2 Proses Analisis <i>Gaussian Kernel Fuzzy C-Means Clustering</i>	43
3.3 Interpretasi Hasil <i>Cluster</i>	50
3.4 Evaluasi Hasil <i>Cluster Gaussian Kernel Fuzzy C-Means</i>	51
BAB IV PENUTUP	55
4.1 Kesimpulan	55
4.2 Saran	55
DAFTAR PUSTAKA	56
LAMPIRAN	59

DAFTAR TABEL

Tabel	Halaman
1. Kriteria Penilaian <i>Clustering</i> Berdasarkan <i>Average Silhouette Value</i>	9
2. Struktur Variabel Data	12
3. Nilai <i>Davies-Bouldin Index</i> Untuk Tiap Jumlah <i>Cluster</i> GKFCM	41
4. Titik Pusat <i>Cluster</i> Pada Iterasi Ke-1	43
5. Matriks Keanggotaan <i>Fuzzy</i> GKFCM Pada Iterasi Ke-1	46
6. Titik Pusat <i>Cluster</i> Pada Iterasi Ke-2	47
7. Titik Pusat <i>Cluster</i> Pada Iterasi Ke-93.....	48
8. Matriks Keanggotaan <i>Fuzzy</i> GKFCM Pada Iterasi Ke-93	48
9. Nilai Rata-rata untuk Masing-masing Kelompok Wilayah.....	51

DAFTAR GAMBAR

Gambar	Halaman
1. Peta <i>Cluster</i> GKFCM Kabupaten/Kota Sulawesi Selatan.....	50
2. Nilai <i>Silhouette Value</i> Hasil <i>Clustering</i> GKFCM pada Data Kriminalitas Kabupaten/Kota di Sulawesi Selatan	50

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Kriminalitas Sulawesi Selatan Tahun 2022	60
2. Nilai <i>Kernel Gaussian</i> Titik Pusat Dengan Data Pada Iterasi Ke-1	67
3. Jarak <i>Euclidian</i> Antar Titik Data Kriminalitas Tiap Kabupaten/Kota.....	35
4. Nilai <i>Silhouette Value</i> Hasil GKFCM pada Keseluruhan Data.....	38

DAFTAR ISTILAH

Istilah	Arti dan Penjelasan
<i>Cluster</i>	Kelompok data yang memiliki kesamaan atau kedekatan dalam suatu dataset tertentu
<i>Fuzzy</i>	Menunjukkan kepemilikan data yang tidak tegas, menggunakan nilai derajat kepemilikan
<i>Outlier</i>	Data yang berbeda jauh dari data lain, seringkali dianggap anomali
Algoritma	Prosedur atau langkah-langkah sistematis untuk menyelesaikan masalah dengan metode tertentu
<i>C-Means</i>	Metode <i>clustering</i> yang menentukan <i>centroid</i> untuk setiap <i>cluster</i> berdasarkan data tertentu.
<i>Machine Learning</i>	Cabang AI yang membuat sistem belajar dari data tanpa eksplisit diprogram.
<i>Centroid</i>	Titik tengah atau representatif dari suatu <i>cluster</i> dalam metode <i>clustering</i> tertentu.
Invers	Operasi matematika untuk menemukan kebalikan suatu matriks, berguna dalam berbagai analisis.
<i>kovarian</i>	Ukuran seberapa dua variabel berubah bersama, menunjukkan hubungan linier antara variabel
Dispersi	Sebaran atau variabilitas data dalam dataset, menunjukkan penyebaran nilai data
<i>Multivariat</i>	Melibatkan banyak variabel atau dimensi dalam analisis data atau model statistik
<i>Index</i>	Indikator atau penunjuk posisi dalam data, berguna untuk mengidentifikasi elemen
<i>Min</i>	Nilai terkecil dalam dataset, menunjukkan batas bawah dari data yang dianalisis
<i>Max</i>	Nilai terbesar dalam dataset, menunjukkan batas atas dari data yang dianalisis
<i>Mean</i>	Rata-rata dari nilai-nilai dalam dataset, memberikan representasi pusat data
Iterasi	Pengulangan langkah-langkah dalam proses algoritma untuk mencapai hasil optimal

DAFTAR LAMBANG/ SINGKATAN

Lambang/Singkatan	Arti dan Penjelasan
m	Degree of <i>fuzzification</i>
n	Jumlah banyaknya data
c	Jumlah banyaknya <i>cluster</i>
x_i	Data ke- i
v_j	Pusat <i>cluster</i> ke- j
u_{ij}	Nilai keanggotaan data ke- i pada <i>cluster</i> ke- j
σ^2	Variansi dari data
$\psi(x_i, v_j)$	Nilai <i>gaussian kernel</i> data x_i dengan pusat <i>cluster</i> v_j
DBI	<i>Davies-bouldin index</i>
MAE_j	Jarak intra- <i>cluster</i> ke- j
d_j	Nilai rata-rata <i>cluster</i> ke- j
T_j	Jumlah data pada <i>cluster</i> ke- j
$R_{j,k}$	Kesamaan antar <i>cluster</i> ke- j dan ke- k
$\ d_j - d_k\ $	Jarak <i>euclidian mean cluster</i> ke- j dan <i>cluster</i> ke- k
$\ x_i - d_j\ $	Jarak <i>euclidian mean cluster</i> ke- j dan data ke- i
ASV	<i>Average silhouette value</i>
$s(i)$	Nilai <i>silhouette</i> masing-masing objek
$a(i)$	Jarak rata-rata sampel i ke sampel lain dalam <i>cluster</i>
$b(i)$	Minimum jarak rata-rata sampel dari sampel i ke <i>cluster</i> lain.
nc	Jumlah data dalam <i>cluster</i> c
nk	Jumlah data di luar <i>cluster</i> c

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis *cluster* adalah analisis untuk mengelompokkan data yang memiliki elemen dan karakteristik serupa ke dalam kelompok (*cluster*) yang berbeda. Analisis *cluster* digunakan untuk mengeksplorasi dan mengelompokkan observasi-observasi yang serupa berdasarkan pola-pola dari kumpulan observasi tersebut. Analisis *cluster* berguna untuk meringkas data dengan cara mengelompokkan data berdasarkan kesamaan karakteristik data yang diteliti (Sitepu dan Gultom, 2011). Tujuan utama analisis *cluster* adalah mengklasifikasi objek (kasus atau elemen) kedalam kelompok-kelompok yang relatif homogen berdasarkan variabel-variabel yang dipertimbangkan dalam penelitian. Analisis *cluster* terbagi atas dua metode, yaitu metode hierarki dan metode *non-hierarki* (Goreti dkk., 2016).

Fuzzy C-Means (FCM) yang merupakan salah satu bagian dari metode *clustering non-hierarki* adalah suatu teknik *clustering* yang mana keanggotaan data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Derajat keanggotaan pada FCM memungkinkan suatu titik data bisa termasuk kedalam lebih dari satu kemungkinan *cluster* (Hardiyanto dkk., 2012). Algoritma FCM memiliki kelebihan dalam penempatan pusat *cluster* yang lebih tepat dibandingkan dengan metode *cluster* lainnya. FCM dapat mengatasi kelemahan dari metode seperti *K-Means* dengan memperbarui pusat *cluster* menggunakan nilai keanggotaan *fuzzy*

setiap data (Nurkholis dkk., 2022). FCM dikenal baik karena dapat mengatasi ketidakpastian dan fleksibilitas keanggotaan observasi dalam mendeteksi keanggotaan *cluster* setiap data (Nidyashofa dan Istiawan, 2017).

Metode FCM menggunakan jarak antar data dalam menentukan nilai keanggotaan observasi, sehingga metode FCM sangat rentan terhadap pengaruh *noise* dan perbedaan skala nilai pada data (Chowdhary dkk., 2020). Solusi yang dapat diterapkan adalah dengan menggunakan *kernel gaussian* pada metode FCM yang membantu dalam memetakan hubungan antar titik data. Metode FCM yang menggunakan *kernel gaussian* dikenal dengan sebutan *Gaussian Kernel Fuzzy*

C-Means (GKFCM). Metode GKFCM bertujuan untuk meningkatkan kemampuan FCM dalam menangkap struktur dan perbedaan variabel yang lebih kompleks dalam data dengan memetakan data menggunakan *kernel* (Liu dkk., 2019). Metode GKFCM memiliki parameter jumlah *cluster* yang perlu untuk dievaluasi untuk memperoleh jumlah *cluster* paling optimal dengan melihat seberapa baik data telah dikelompokkan oleh masing-masing jumlah *cluster* yang ditentukan. Davies Bouldin Index (DBI) merupakan salah satu metrik yang umum digunakan untuk evaluasi hasil *cluster* (Fikri dkk., 2023). DBI mengukur kualitas *cluster* dengan cara

membandingkan jarak antara *cluster* dan jarak dalam *cluster*. Semakin rendah nilai DBI, semakin baik kualitas *clustering* yang dihasilkan, menunjukkan bahwa *cluster* lebih seragam dan terpisah baik dari data pada *cluster* berbeda (Fajar dkk., 2024). DBI dapat digunakan untuk mengevaluasi seberapa baik pengelompokan yang dihasilkan oleh metode GKFCM.

Penelitian yang dilakukan oleh Syam tahun 2021 menggunakan metode FCM dalam pengelompokan kabupaten/kota Sulawesi Selatan berdasarkan jumlah tenaga kesehatan menghasilkan 6 *cluster* dengan nilai *pseudo F-statistic* sebesar 16,1843 namun kota Makassar dan Bone dikategorikan sebagai *outlier*. Penelitian lainnya dilakukan oleh Chang-Chien dkk. tahun 2021 menggunakan FCM dan GKFCM pada beberapa data berbeda, hasil dari penelitian ini menunjukkan bahwa metode GKFCM memperoleh nilai *average error* sebesar 0,29 lebih baik dibandingkan FCM dengan *average error* sebesar 0,34. Gosain dan Dahiya tahun 2020 juga telah melakukan penelitian pada data sintesis yang mengandung *outlier*, memperoleh hasil nilai *average error* sebesar 65,32 pada FCM lalu 6,2 pada FCM berbasis intuisi dan *gaussian kernel*, menunjukkan bahwa *gaussian kernel* membantu model menangani pengaruh dari data *outlier* pada proses *clustering*.

Berdasarkan uraian latar belakang yang telah dipaparkan serta penelitian sebelumnya, menunjukkan bahwa metode FCM mampu melakukan klasifikasi pada data dengan kompleksitas tertentu namun metode FCM sangat sensitif terhadap *outlier* dan *noise*. Pengembangan metode FCM yaitu GKFCM menjadi solusi pada kelemahan FCM terhadap *outlier* dan *noise* menggunakan pendekatan *kernel*, sehingga pada penelitian tugas akhir ini akan dilakukan pengelompokan kabupaten/kota di Sulawesi Selatan menggunakan metode GKFCM pada data kriminalitas setiap kabupaten/kota di Sulawesi Selatan.

1.2 Rumusan Masalah

Rumusan masalah dari penulisan tugas akhir ini adalah sebagai berikut:

1. Bagaimana menentukan jumlah *cluster* terbaik menggunakan nilai *Deviasi Bouldin Index* serta menentukan karakteristik setiap kelompok kabupaten/kota di Provinsi Sulawesi Selatan yang terbentuk berdasarkan data kriminalitas menggunakan metode *Gaussian Kernel Fuzzy C-Means*?

1.3 Batasan Masalah

Batasan Fokus penelitian ini akan dibatasi pada data kriminalitas tahun 2022 Provinsi Sulawesi Selatan. yang diperoleh dari buku Publikasi BPS "Provinsi Sulawesi Selatan Dalam Angka Tahun 2023". Indikator yang digunakan dalam penelitian ini adalah Jumlah Kejahatan yang Dilaporkan, Risiko Penduduk Terkena Kejahatan per 100.000 Penduduk, Persentase Penyelesaian Kejahatan dan Selang Waktu Terjadinya Tindak Kejahatan.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh jumlah *cluster* terbaik menggunakan nilai *Deviasi Bouldin Index* serta mendapatkan karakteristik setiap kelompok kabupaten/kota di Provinsi Sulawesi Selatan yang terbentuk berdasarkan data kriminalitas menggunakan metode *Gaussian Kernel Fuzzy C-Means*.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat membantu dalam memahami lebih baik bagaimana penduduk di kabupaten/kota di Provinsi Sulawesi Selatan terbagi berdasarkan karakteristik tindak kriminalitas. Menggunakan metode *Gaussian Kernel Fuzzy C-Means*, penelitian ini dapat memberikan pengelompokan yang lebih akurat dari metode tradisional. Hasil penelitian ini akan membantu pemerintah dalam mengidentifikasi *cluster* populasi yang memiliki karakteristik yang serupa.

1.6 Landasan Teori

1.6.1 Analisis Cluster

Analisis *cluster* merupakan salah satu metode data *mining* yang bersifat unsupervised yang memiliki arti bahwa karakteristik tiap *cluster* tidak ditentukan sebelumnya melainkan berdasarkan kemiripan atribut-atribut dari suatu kelompok atau *cluster*. Analisis *cluster* membagi data menjadi kelompok-kelompok atau *cluster* berdasarkan suatu kemiripan atribut-atribut diantara sekumpulan data, atribut-atribut yang mirip tersebut dipresentasikan sebagai titik-titik dalam ruang multidimensi (Andini dan Farokhah, 2022).

Analisis *cluster* merupakan metode multivariat yang mempunyai tujuan untuk mengelompokkan objek berdasarkan kesamaan karakteristik. Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama (Ferezagia, 2018). Pembuatan *cluster* pada wilayah sebuah negara memungkinkan pemerintah atau pengurus wilayah untuk lebih memahami kinerjanya dalam mengelola aspek atau variabel yang mempengaruhi kehidupan masyarakat, sehingga pemerintah dapat menentukan beberapa strategi yang efektif dalam rangka meningkatkan kualitas kehidupan masyarakat daerah tersebut (Handoyono, 2022).

1.6.2 Fuzzy C-Means Clustering

Salah satu algoritma *fuzzy clustering* yang paling banyak digunakan adalah Algoritma *Fuzzy C-Means* (FCM). Algoritma FCM didasarkan pada konsep *fuzzy*,

dimana *fuzzy* merupakan suatu nilai yang dapat bernilai benar atau salah secara bersamaan, namun seberapa besar nilai kebenaran dan kesalahannya tergantung pada derajat keanggotaan yang dimilikinya (Sovina dan Harahap, 2022). Sistem logika *fuzzy* digunakan dalam sebuah sistem yang dibangun pada sebuah fenomena yang bersifat samar-samar. Logika *fuzzy* dapat dianggap sebagai pendekatan untuk memetakan suatu ruang *input* atau masukan ke dalam suatu ruang *output* atau keluaran (Setiawan dkk., 2018). Dasar dari logika *fuzzy* adalah teori himpunan *fuzzy*. Pada teori himpunan *fuzzy*, peranan derajat keanggotaan sebagai penentu keberadaan elemen dalam suatu himpunan sangatlah penting. Nilai keanggotaan biasa disebut derajat keanggotaan atau membership value menjadi ciri utama dari penalaran dengan logika *fuzzy* (Zahida, 2018).

Algoritma FCM mencoba mengelompokkan kumpulan terbatas n elemen $X = (x_1, x_2, \dots, x_n)$ kedalam c kelompok data atau *cluster* dengan menggunakan nilai *fuzzy* yang didasarkan pada jarak data ke pusat *cluster* $V = (v_1, v_2, \dots, v_c)$ dengan memperhatikan beberapa variabel tertentu. Algoritma FCM memberikan daftar pusat *cluster* c dan matriks partisi U dengan nilai u_{ij} menunjukkan kemungkinan elemen x_i termasuk dalam *cluster* y_j (Vignesh, 2019). Matriks partisi U merupakan matriks berukuran $n \times c$, $U = [u_{ij}]$ yang merupakan matriks partisi untuk menentukan nilai dari $\{y_j\}$ atau nilai sebuah data termasuk kedalam kumpulan data dengan ketentuan sebagai berikut (Bezdek dkk., 1988):

$$u_j(x_i) = u_{ij} = \begin{cases} 1; & x_i \in y_j \\ 0 & \text{lainnya} \end{cases}$$

$$\sum_{i=1}^n u_{ij} > 0 \quad \text{untuk semua } i;$$

$$\sum_{j=1}^c u_{ij} = 1 \quad \text{untuk semua } j;$$

Algoritma FCM bertujuan untuk meminimalkan fungsi tujuan yang memiliki derajat *fuzzifier* m . *Fuzzifier* m menentukan tingkat ketidakjelasan *cluster* yang pada umumnya nilai m diatur pada nilai $m = 2$. Ketika m mendekati nilai 1, pusat *cluster* yang paling dekat dengan observasi diberikan nilai yang jauh lebih besar dibandingkan yang lain (Kononenko dan Matjazkucar, 2007). Algoritma FCM memberikan derajat kepemilikan terhadap *cluster* pada setiap titik data, seperti dalam logika *fuzzy*, sehingga setiap titik bukan sepenuhnya hanya terdapat pada satu *cluster*. Titik pengamatan di tepi *cluster* mungkin memiliki tingkat derajat kepemilikan *cluster* yang lebih rendah dibandingkan titik yang lebih dekat dengan pusat *cluster* (Vignesh, 2019).

Algoritma FCM *Clustering* bertujuan untuk membentuk kelompok data yang padat dengan cara meminimalkan fungsi objektifnya. Fungsi objektif ini melibatkan penjumlahan perbedaan berpasangan antara titik data dan pusat *cluster*, dengan

tingkat keanggotaan untuk setiap *cluster* pada matriks \mathbf{U} dikalikan dengan kuadrat jarak antara titik data dan pusat *cluster* \mathbf{V} . Proses meminimalkan fungsi pada FCM bertujuan untuk mencari tingkat keanggotaan yang menghasilkan kelompok yang kompak di sekitar pusat data yang signifikan, memungkinkan elemen data untuk termasuk dalam lebih dari satu kelompok dengan tingkat keanggotaan yang dapat diukur secara *fuzzy*. Fungsi objektif FCM ditampilkan pada persamaan (Askari, 2020):

$$J_m(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m (\|x_i - v_j\|)^2; \quad 1 \leq m \leq \infty \quad (1)$$

Keterangan:

- u_{ij} : Nilai keanggotaan data ke- i pada *cluster* ke- j
- v_j : Pusat *cluster* ke- j
- x_i : Data ke- i
- m : Tingkat ketidakpastian (*degree of fuzziness*)
- n : Jumlah banyaknya data
- c : Jumlah banyaknya *cluster*

Nilai u_{ij} merupakan nilai fungsi keanggotaan dari data yang memiliki nilai diantara 0 dan 1. Persamaan untuk menghitung nilai u_{ij} untuk setiap data x_i pada *cluster* c_j adalah sebagai berikut (Alam dkk., 2019):

$$u_{ij} = \sum_{k=1}^c \left(\frac{(\|x_i - v_j\|)}{(\|x_i - v_k\|)} \right)^{\left(-\frac{2}{m-1}\right)} \quad (2)$$

Dengan

$$v_j = \frac{\sum_{i=1}^n (u_{ij})^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (3)$$

$$\|x_i - v_i\|^2 = \sqrt{\sum (x_i - v_i)^2} \quad (4)$$

Pada persamaan diatas $\|x_i - v_j\|$ dapat direpresentasikan sebagai jarak *Euclidean*. Algoritma *Fuzzy C-Means* (FCM) mengoptimalkan fungsi objektif $J_m(\mathbf{U}, \mathbf{V})$ melalui pembaruan berkelanjutan dari \mathbf{U} dan \mathbf{V} . Iterasi dilakukan hingga perbedaan antara tingkat keanggotaan pada dua iterasi berturut-turut $\|U_q - U_{q-1}\| \leq e$, kurang dari atau sama dengan nilai toleransi e , dengan e berada dalam rentang 0 hingga 1, proses ini berlangsung hingga batas toleransi tercapai. (Alam dkk., 2019).

1.6.3 Gaussian Kernel

Gaussian kernel adalah fungsi matematika yang mengukur kesamaan antara dua titik data berdasarkan jarak *euclidean* antara titik-titik tersebut. *Gaussian kernel* digunakan untuk menentukan seberapa mirip dua titik data berdasarkan titik data pada ruang fitur, nilai fungsi *gaussian kernel* berkurang secara eksponensial seiring dengan bertambahnya jarak antara titik-titik tersebut, memungkinkan *gaussian kernel* ini untuk menangkap hubungan non-linear yang kompleks dalam data (Kowalczyk, 2017). Berikut merupakan persamaan fungsi *gaussian kernel* untuk melihat hubungan antara dua titik data (Jiang dkk., 2020)

$$\psi(x_i, v_j) = \exp\left(\frac{-\|x_i - v_j\|^2}{\sigma^2}\right) \quad (5)$$

Keterangan:

- u_{ij} : Nilai keanggotaan data ke- i pada *cluster* ke- j
- v_j : Pusat *cluster* ke- j
- x_i : Data ke- i
- $\|x_i - v_j\|$: Jarak *euclidian* data x_i dari pusat *cluster* v_j
- σ^2 : Varians dari data

Ketika data terdistribusi tidak merata atau beberapa bagian dari data sangat padat dengan banyak titik data yang berdekatan, sementara bagian lain dari data sangat jarang dengan titik data yang tersebar jauh. Situasi ini sering kali muncul dalam *dataset* nyata di mana pola atau *cluster* data tidak selalu linier atau seragam. *gaussian kernel*, dirancang untuk menangani kompleksitas ini dengan sangat baik. Berikut adalah beberapa alasan mengapa RBF *kernel* penting untuk data yang *gaussian kernel* dapat memetakan data ke ruang fitur yang lebih tinggi di mana pola *non-linear* menjadi lebih mudah diidentifikasi dan dipisahkan. Ini memungkinkan model untuk menangkap kompleksitas dan struktur yang ada dalam data yang tidak terdistribusi secara merata (Rochim et al., 2021).

1.6.4 Gaussian Kernel Fuzzy C-Means Clustering

Fungsi objektif standar dari algoritma *Fuzzy C-Means* (FCM) bertujuan untuk membagi himpunan data X menjadi V kelompok atau *cluster* dengan cara meminimalkan suatu fungsi matematis tertentu, yang disebut sebagai fungsi objektif. Fungsi objektif FCM ditampilkan pada persamaan (1). Tahapan proses *clustering* GKFCM adalah dengan meminimalkan fungsi objektif pada persamaan dapat dituliskan sebagai berikut:

$$J_m(\mathbf{U}, \mathbf{V}) = 2 \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m (1 - \psi(x_i, v_j)) \quad (6)$$

Dengan batasan

$$\sum_{i=1}^c u_{ij} = 1$$

Keterangan:

- $\psi(x_i, v_j)$: Nilai *gaussian kernel* data x_i dengan pusat *cluster* v_j
 u_{ij} : Nilai keanggotaan data ke- i pada *cluster* ke- j
 v_j : Pusat *cluster* ke- j
 x_i : Data ke- i
 m : Tingkat ketidakpastian (*degree of fuzziness*)
 n : Jumlah banyaknya data
 c : Jumlah banyaknya *cluster*

Fungsi *gaussian* $\psi(x_i, v_k)$ yang dimasukkan kedalam persamaan untuk mengevaluasi jarak merupakan fungsi *gaussian* pada Persamaan (6). Persamaan yang digunakan pada tiap iterasi untuk memperbarui nilai matriks partisi dan pusat *cluster* sehingga dapat meminimalkan fungsi objektif sebagai berikut:

$$u_{ij} = \frac{\left((1 - \psi(x_i, v_j))^{-1} \right)^{\frac{1}{m-1}}}{\sum_{k=1}^c \left((1 - \psi(x_i, v_k))^{-1} \right)^{\frac{1}{m-1}}} \quad (7)$$

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m \psi(x_i, v_j) x_i}{\sum_{i=1}^n u_{ij}^m \psi(x_i, v_j)} \quad (8)$$

Dengan menggunakan fungsi *gaussian* ini, kita dapat mengevaluasi jarak antara data point (x_i) dan pusat *cluster* (v_j) dengan memasukkan nilai-nilai tersebut ke dalam fungsi *gaussian*. Ini memberikan representasi jarak yang mempertimbangkan distribusi *gaussian* antara *data point* dan pusat *cluster*, yang sesuai dengan pendekatan *gaussian* yang diadopsi (Jiang dkk., 2020).

1.6.5 Davies-Bouldin Index

David L. Davies dan Donald W. Bouldin memperkenalkan sebuah metode dan memberikan nama pada metode tersebut dengan nama kedua beliau, yaitu *Davies-Bouldin Index* (DBI) yang digunakan untuk mengevaluasi *cluster*. Evaluasi menggunakan DBI memiliki skema evaluasi *cluster* internal, di mana hasil *cluster* yang baik atau tidak dilihat dari kuantitas dan kedekatan antara hasil *cluster*. DBI adalah satu metode yang digunakan untuk mengukur validitas *cluster* dalam suatu metode pengelompokan, di mana hubungan antar data dalam kelompok didefinisikan sebagai jumlah kedekatan data ke titik pusat *cluster* diikuti oleh pemisahan berdasarkan jarak antara titik pusat *cluster* ke titik pusat *cluster* lainnya. Pengukuran dengan DBI memaksimalkan jarak antar *cluster* misalnya jarak antara *cluster* C_i dan C_j dan pada saat yang sama mencoba untuk meminimalkan jarak antara titik-titik

dalam suatu *cluster*. Jika jarak antar *cluster* maksimal, itu berarti bahwa kesamaan karakteristik antara setiap *cluster* bernilai kecil sehingga perbedaan antara *cluster* tampak lebih jelas. Jika jarak antar data dalam *cluster* minimum atau data dalam *cluster* yang sama saling berdekatan, itu berarti bahwa setiap objek dalam *cluster* memiliki tingkat kesamaan karakteristik yang tinggi (Mughnyanti dkk., 2020).

Prinsip pendekatan pengukuran DBI adalah memaksimalkan jarak antar *cluster* serta meminimalkan jarak dalam *cluster*. Semakin kecil nilai DBI menunjukkan skema *cluster* yang paling optimal. Perhitungan DBI dijabarkan dalam persamaan berikut (Gie dan Jollyta, 2020):

$$DBI = \frac{1}{c} \sum_{j=1}^c \max_{j \neq k} R_{j,k} \quad (9)$$

Dengan

$$R_{j,k} = \frac{MAE_j + MAE_k}{\|d_j - d_k\|} \quad (10)$$

$$MAE_j = \left(\frac{1}{T_j} \sum_{x \in C_j} \|x_i - d_j\| \right) \quad (11)$$

Keterangan

DBI	: <i>Davies-bouldin index</i>
MAE_j	: Jarak intra- <i>cluster</i> ke- j
d_j	: Nilai rata-rata <i>cluster</i> ke- j
T_j	: Jumlah data pada <i>cluster</i> ke- j
$R_{j,k}$: Kesamaan antar <i>cluster</i> ke- j dan ke- k
c	: Jumlah <i>cluster</i>
$\ d_j - d_k\ $: Jarak <i>euclidian mean cluster</i> ke- j dan <i>cluster</i> ke- k
$\ x_i - d_j\ $: Jarak <i>euclidian mean cluster</i> ke- j dan data ke- i

1.6.6 Average Silhouette Value

Average Silhouette Value (ASV) diperkenalkan oleh Peter J. Rousseeuw pada tahun 1986, adalah metode yang digunakan untuk memvalidasi data *cluster*. ASV memberikan representasi yang menunjukkan seberapa baik setiap titik data sesuai dengan *cluster* nya (Subbalakshmi dkk., 2015). ASV adalah suatu nilai yang mampu mengukur tingkat kehomogenan anggota dalam internal *cluster* dan tingkat keheterogenan anggota antar *cluster*. Nilai ASV evaluasi *cluster* berada pada interval -1 hingga 1 (Septianingsih, 2022). Nilai *silhouette value* mendekati 1,

menunjukkan bahwa titik data secara signifikan lebih dekat dengan titik data dalam *cluster* yang sama dibandingkan dengan titik data *cluster* lain. Ketika nilai mendekati 0, menandakan adanya ketidakpastian dalam penempatan titik data tersebut, dengan kemungkinan bahwa data tersebut bisa masuk ke dalam *cluster* lain yang memiliki karakteristik serupa. Sementara itu, nilai *silhouette value* yang mendekati -1 mengindikasikan penempatan yang salah, dengan titik data lebih dekat dengan *cluster* lain daripada yang seharusnya, mengindikasikan ketidakcocokan antara titik data dan *cluster*nya (Shahapure dan Nicholas, 2020). Berikut merupakan Tabel 2.1 yang menampilkan standar kriteria evaluasi *cluster* menggunakan nilai ASV menurut (Guntara dan Suprawoto, 2022):

Tabel 1. Kriteria Penilaian *Clustering* Berdasarkan *Average Silhouette Value*

Average Silhouette Value	Kriteria
$0,7 < ASV \leq 1$	Struktur <i>cluster</i> yang sangat baik
$0,5 < ASV \leq 0,7$	Struktur <i>cluster</i> yang cukup baik
$0,25 < ASV \leq 0,5$	Struktur <i>cluster</i> yang lemah
$ASV \leq 0,25$	Tidak ada struktur <i>cluster</i>

Nilai ASV dapat dihitung setelah setiap data dikelompokan dan nilai tengah masing-masing *cluster* telah dihitung. Persamaan untuk menghitung ASV adalah sebagai berikut (Septianingsih, 2022):

$$ASV = \frac{1}{N} \sum_{i=1}^N s(i) \quad (12)$$

dengan:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (13)$$

$$a(i) = \frac{1}{nc - 1} \sum_{r=1}^{nc-1} d(x_{ic} - x_{rc}) \quad (14)$$

$$d(i, C) = \frac{1}{nk} \sum_{r=1}^{nk} d(x_i - x_{rc}) \quad (15)$$

$$b(i) = \min d(i, C) \quad (16)$$

keterangan:

- ASV : *Average silhouette value*
 $s(i)$: Nilai *silhouette* masing-masing objek
 $a(i)$: Jarak rata-rata sampel i ke sampel lain dalam *cluster*
 $b(i)$: Minimum jarak rata-rata sampel dari sampel i ke *cluster* lain.

- nc : Jumlah data dalam *cluster c*
 nk : Jumlah data di luar *cluster c*
 $d(x_{ic} - x_{rc})$: jarak data ke - i dengan data ke - r dalam satu *cluster*
 $d(x_i - x_{rc})$: jarak data ke - i dengan data ke - r tidak satu *cluster*

1.6.7 Kriminalitas

Tindak kriminalitas adalah suatu tindakan yang melanggar hukum, norma, dan peraturan sosial yang berlaku dalam suatu kelompok masyarakat, sehingga bagi yang melakukan pelanggaran harus diberikan sanksi atau penghukuman. Tindak kriminalitas sangat berdampak negatif terhadap kesejahteraan masyarakat di sebuah wilayah, karena dapat menimbulkan rasa tidak aman, kecemasan, kepanikan, dan ketakutan. Tingginya kasus kriminalitas yang terjadi disuatu wilayah dapat menggambarkan besarnya risiko setiap masyarakat menjadi korban kriminalitas, risiko tersebut dapat dinyatakan dalam parameter tingkat kejahatan atau kriminalitas (Sulistiani, 2022). Secara umum kriminalitas dapat dijelaskan dengan indikator sebagai berikut:

1. Jumlah Kejadian yang Dilaporkan

Jumlah kejadian yang dilaporkan adalah setiap peristiwa yang dilaporkan masyarakat pada Polri, atau peristiwa dimana pelakunya tertangkap tangan oleh kepolisian. Laporan masyarakat ini akan dicatat dan ditindaklanjuti oleh Polri jika dikategorikan memiliki cukup bukti. Kejadian kejahatan yang dilaporkan terdiri atas kejahatan terhadap nyawa, fisik/badan, kesusilaan, hak milik/barang, narkoba, ketertiban umum, serta penipuan, penggelapan, dan korupsi (BPS, 2016).

2. Risiko penduduk Terkena Kejahatan per 100.000 Penduduk (*Crime Rate*)

Crime rate merupakan angka yang dapat menunjukkan tingkat kerawanan terjadinya tindak kejahatan pada suatu wilayah tertentu dalam waktu tertentu. Semakin tinggi angka *crime rate* maka tingkat kerawanan akan kejahatan suatu daerah semakin tinggi pula, dan sebaliknya. Selama periode tahun 2018–2020, tingkat resiko terkena tindak kejahatan (*crime rate*) setiap 100.000 penduduk mengalami penurunan (BPS, 2021). Perhitungan Risiko penduduk Terkena Kejahatan per 100.000 Penduduk adalah sebagai berikut (Kepala Kepolisian Negara Republik Indonesia, 2009):

$$crime\ rate = \frac{Jumlah\ Kejadian\ yang\ Dilaporkan}{Jumlah\ Penduduk} \times 100.000$$

3. Persentase Penyelesaian Tindak Pidana (*Clearance Rate*)

Clearance rate adalah proporsi kejahatan yang diketahui polisi yang diselesaikan (penangkapan dilakukan) versus semua kejahatan yang diketahui polisi. Dalam peradilan pidana, tingkat penyelesaian dihitung dengan membagi jumlah kejahatan yang "dibersihkan" (dikenakan dakwaan) dengan jumlah total kejahatan yang tercatat. Tingkat penyelesaian digunakan oleh berbagai kelompok sebagai ukuran

kejahatan yang diselesaikan oleh polisi. Tingginya presentase *clearance rate* oleh pihak kepolisian belum mampu menekan angka kejahatan. Ketika angka *clearance rate* meningkat hal tersebut menunjukkan tingkat keberhasilan dari pihak kepolisian dalam menyelesaikan kasus yang dilaporkan (BPS, 2016). Perhitungan Persentase Penyelesaian Tindak Pidana adalah sebagai berikut (Kepala Kepolisian Negara Republik Indonesia, 2009):

$$Clearance Rate = \frac{Kasus\ yang\ diselesaikan}{Jumlah\ Kejadian\ yang\ Dilaporkan} \times 100\%$$

4. Selang Waktu Terjadinya Tindak Pidana (*Crime Clock*)

Crime clock menunjukkan pada beberapa kali kejahatan yang terjadi pada setiap jam. Selang waktu terjadinya kejahatan di Indonesia selama 2 menit 11 detik pada 2021. Semakin tinggi *crime clock* maka semakin aman pula daerah tersebut dari tindak kejahatan (BPS, 2021). Perhitungan Selang Waktu Terjadinya Tindak Pidana adalah sebagai berikut (Kepala Kepolisian Negara Republik Indonesia, 2009):

$$Crime\ Clock = \frac{365 \times 24}{Jumlah\ Kejadian\ yang\ Dilaporkan}$$

BAB II METODOLOGI PENELITIAN

2.1 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan yaitu data data Kriminalitas Sulawesi Selatan tahun 2022 pada publikasi Provinsi Sulawesi Selatan Dalam Angka 2023 yang dapat diakses melalui situs resmi BPS Provinsi Sulawesi Selatan pada tautan <https://sulsel.bps.go.id>.

2.2 Struktur Data

Struktur data yang digunakan berupa data *cross section* dengan 24 observasi yang meliputi 24 kabupaten/kota di Sulawesi Selatan. Keseluruhan data yang digunakan terdapat pada Lampiran 1. Adapun variabel-variabel yang digunakan pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2. Struktur Variabel Data

Variabel	Keterangan	Satuan
X1	Jumlah Kejahatan yang Dilaporkan	Kasus
X2	Risiko Penduduk Terkena Kejahatan per 100.000 Penduduk	Penduduk
X3	Persentase Penyelesaian Kejahatan	Persen
X4	Selang Waktu Terjadinya Kejahatan	Jam

2.3 Metode Analisis

Pendekatan analisis yang digunakan dalam penelitian ini adalah Pengelompokan dengan metode GKFCM. Pengolahan data dilakukan dengan bantuan *Software Jupyter Notebook*. Tahapan analisis data yang dilakukan dalam penelitian ini yaitu sebagai berikut:

1. Menentukan jumlah *cluster* terbaik menggunakan nilai *Davies-Bouldin Index* yang diperoleh dengan menggunakan Persamaan (9) yang membutuhkan nilai pusat dan pengelompokan *cluster* dari tiap observasi. Penentuan pusat *cluster* dan pengelompokan *cluster* tiap observasi dilakukan sebagai berikut:
 - a. Menentukan jumlah *cluster* yang akan digunakan dalam proses *clustering*.
 - b. Menentukan nilai titik pusat *cluster* dengan menggunakan nilai acak sebagai inisiasi awal pusat *cluster* untuk iterasi pertama dan Persamaan (8) pada iterasi kedua dan seterusnya.
 - c. Menghitung nilai *kernel* setiap titik observasi dengan setiap titik pusat *cluster* menggunakan Persamaan (6) sehingga diperoleh matriks berisi nilai

- kernel gaussian* antara setiap titik pusat *cluster* dengan setiap observasi.
- d. Nilai *kernel* dan nilai pusat *cluster* yang telah diperoleh digunakan untuk memperbarui nilai matriks keanggotaan *fuzzy U* dengan Persamaan (7). Nilai matriks keanggotaan *fuzzy* yang telah diperbarui digunakan untuk proses iterasi selanjutnya.
 - e. Langkah b hingga d terus diiterasikan hingga batas iterasi telah tercapai atau nilai $\|U_q - U_{q-1}\| \leq e$.
 - f. Menghitung nilai *Davies-Bouldin Index* menggunakan Persamaan (9), lalu mengulangi seluruh proses dari langkah a hingga b menggunakan nilai *cluster* yang berbeda
2. Mengevaluasi kualitas *cluster* hasil pengelompokan GKFCM menggunakan nilai ASV.
 3. Setelah memperoleh jumlah *cluster* terbaik, maka selanjutnya pengelompokan yang telah diperoleh digunakan untuk menggambarkan distribusi pengelompokan kabupaten/kota berdasarkan data kriminalitas pada peta Sulawesi Selatan.
 4. Hasil yang diperoleh diinterpretasikan lalu dibuatkan kesimpulan.