

**PENENTUAN *CLUSTER* TERBAIK MENGGUNAKAN METODE *FUZZY*
C-MEANS DENGAN ANALISIS KOMPONEN UTAMA**

**ANDI MUHAMMAD ALIF REZA BAHRI
H051201037**



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

PENENTUAN *CLUSTER* TERBAIK MENGGUNAKAN METODE *FUZZY C-MEANS* DENGAN ANALISIS KOMPONEN UTAMA

ANDI MUHAMMAD ALIF REZA BAHRI
H051201037



Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Statistika

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

**PENENTUAN CLUSTER TERBAIK MENGGUNAKAN METODE
FUZZY C-MEANS DENGAN ANALISIS KOMPONEN UTAMA**

Yang disusun dan diajukan oleh

ANDI MUHAMMAD ALIF REZA BAHRI

H051201037

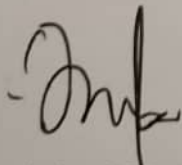
Skripsi,

Telah dipertahankan di depan Panitia Ujian Sarjana Statistika pada 13
Agustus 2024 dan dinyatakan telah memenuhi syarat kelulusan

pada

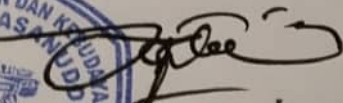
Program Studi Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

Mengesahkan:
Pembimbing tugas akhir,



Anisa, S.Si., M.Si.
NIP. 1973022719988022001

Mengetahui:
Ketua Program Studi,



Dr. Anna Slamiyati, S.Si., M.Si.
NIP. 197708082005012002

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Penentuan *Cluster* Terbaik Menggunakan Metode *Fuzzy C-Means* dengan Analisis Komponen Utama" adalah benar karya saya dengan arahan dari pembimbing Anisa, S.Si., M.Si. 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 ketentuan aturan yang berlaku.

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

Makassar, 13 Agustus 2024



Andi Muhammad Alif Reza Bahri
NIM H051201037

UCAPAN TERIMA KASIH

Puji dan syukur saya panjatkan kepada Allah SWT, Tuhan Yang Maha Esa, atas segala rahmat, hidayah, dan pertolongan-Nya sehingga penelitian yang saya lakukan dapat terlaksana dengan baik. Tidak lupa shalawat dan salam saya haturkan kepada Nabi Muhammad SAW yang menjadi suri tauladan kita sepanjang zaman.

Dengan penuh rasa hormat, saya sangat berterima kasih kepada dosen pembimbing, Ibu **Anisa, S.Si., M.Si.** yang telah memberikan bimbingan, arahan, dan dukungan yang sangat berharga. Kesabaran dan ilmu yang telah disampaikan sangat membantu saya dalam menyelesaikan tugas akhir ini. Terima kasih juga saya sampaikan kepada dosen penguji, Bapak **Dr. Nirwan, M.Si.** dan Ibu **Sitti Sahrinan, S.Si., M.Si.** yang telah meluangkan waktu, perhatian, dan masukan yang diberikan dalam membangun dalam penyempurnaan tugas akhir ini.

Saya mengucapkan terima kasih kepada Ketua Departemen, Ibu **Dr. Anna Islamiyati S.Si., M.Si.** atas dukungan, bimbingan, dan kepemimpinan yang telah memfasilitasi lingkungan akademik yang kondusif. Terima kasih juga saya sampaikan kepada dosen-dosen Departemen Statistika Unhas atas ilmu, pengalaman, dan dedikasi yang telah dibagikan selama perkuliahan. Setiap pelajaran dan wawasan yang diberikan sangat berkontribusi terhadap pencapaian akademik saya. Tidak lupa, saya mengucapkan terima kasih kepada seluruh staf yang bertugas atas bantuan administratif dan dukungan yang diberikan.

Saya ingin menyampaikan rasa terima kasih yang mendalam kepada kedua orang tua tercinta, Bapak **Andi Bahri Bong S.Ag.** dan Ibu **Nurmin Abdurrahman S.E.** atas doa, pengorbanan, dan dukungan tanpa henti yang telah mereka berikan selama saya menempuh pendidikan. Penghargaan yang tulus saya persembahkan kepada mereka. Ucapan terima kasih juga saya sampaikan kepada adik saya, **Andi Muhammad Khaleq Nugraha Bahri** dan seluruh keluarga besar atas kasih sayang dan dukungan yang tak ternilai harganya.

Saya mengucapkan terima kasih kepada sahabat-sahabat dari grup Random (**Rais, Fadlan, Dzaky, Bahar, Mukhlis, Fahmi, Laurine, Aisyah, dan Krisna**), Among (**Nahla dkk.**) dan Statistika 2020 (**Hakam, Theo, dkk.**) yang telah berproses bersama saya, baik dalam suka maupun duka, selama perkuliahan dan penyelesaian tugas akhir ini. Tak lupa, saya ingin mengucapkan terima kasih kepada sahabat terkasih sejak SMA (**Ishak, Naufal, Farhat, Tenri, Ceye, Fatim, dan Eja**) serta sahabat terkasih sejak SMP (**Indra dan Akbar**) yang telah memberikan banyak dukungan dan hiburan yang sangat berharga.

Akhirnya, kepada semua pihak yang telah memberikan bantuan baik secara langsung maupun tidak langsung, saya haturkan terima kasih yang sebesar-besarnya. Semoga segala kebaikan dan dukungan yang telah diberikan mendapatkan balasan yang setimpal dari Allah SWT.

Makassar, 13 Agustus 2024



Andi Muhammad Alif Reza Bahri
NIM H051201037

ABSTRAK

Andi Muhammad Alif Reza Bahri. **Penentuan *Cluster* Terbaik Menggunakan Metode *Fuzzy C-Means* dengan Analisis Komponen Utama** (dibimbing oleh Anisa).

Latar Belakang. Analisis *Cluster* merupakan metode pengelompokan objek ke dalam suatu *cluster* sehingga objek yang berada dalam satu kelompok akan memiliki karakteristik yang mirip dibandingkan dengan objek pada *cluster* lainnya. Salah satu metode yang bisa digunakan dalam mengelompokkan objek data adalah *Fuzzy C-Means* (FCM) yang merupakan pengelompokan non-hierarki yang ditentukan oleh nilai derajat keanggotaan. Salah satu masalah yang mungkin muncul adalah adanya multikolinearitas pada data sehingga perlu diatasi, salah satunya dengan menggunakan Analisis Komponen Utama (AKU). **Tujuan.** Mengelompokkan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan data kriminalitas tahun 2022 dan menentukan cluster terbaik menggunakan metode FCM dengan AKU dan melihat nilai *Davies Bouldin's Index* (DBI) yang lebih kecil. **Metode.** Pengelompokan dengan menggunakan metode FCM dilakukan setelah terlebih dahulu dilakukan AKU pada data untuk mengatasi multikolinearitas. Berdasarkan hasil dari AKU digunakan total keragaman sebesar 80% untuk menggambarkan keragaman dari data. Selanjutnya, data dikelompokkan menggunakan FCM dan dievaluasi dengan DBI. **Hasil.** Berdasarkan total keragaman 80% dari hasil AKU diperoleh empat komponen utama pertama yang signifikan menggambarkan keragaman dari data. Dari empat komponen utama tersebut dibentuk dua *cluster* dan tiga *cluster* dengan metode FCM. Berdasarkan nilai DBI diperoleh bahwa dua *cluster* adalah pengelompokan terbaik dengan nilai DBI sebesar 0,088 dibandingkan dengan tiga *cluster* (nilai DBI 1,226). Hasil pengelompokan FCM tanpa AKU dengan mengeluarkan Kota Makassar dari analisis data diperoleh *cluster* terbaik sebanyak dua *cluster* dengan nilai DBI 0,340. **Kesimpulan.** Pengelompokan FCM dengan AKU dan tanpa AKU pada data kriminalitas menunjukkan bahwa *cluster* terbaik berdasarkan nilai DBI terkecil adalah sebanyak 2 *cluster*.

Kata Kunci: Analisis Komponen Utama, *Davies Bouldin's Index*, *Fuzzy C-Means*, Multikolinearitas.

ABSTRACT

Andi Muhammad Alif Reza Bahri. **Determining the Best Cluster Using the Fuzzy C-Means Method with Principal Component Analysis** (Supervised by Anisa).

Introduction. Cluster analysis is a method for grouping objects into clusters so that objects within the same cluster have similar characteristics compared to those in other clusters. One method that can be used for clustering data objects is Fuzzy C-Means (FCM), a non-hierarchical clustering technique determined by membership degree values. One issue that may arise is multicollinearity in the data, which needs to be addressed, with one solution being the use of Principal Component Analysis (PCA). **Purpose.** The objective is To group the regencies/cities in South Sulawesi Province based on 2022 crime data, identify the best cluster using the FCM method with PCA, and determine the cluster with the smallest Davies Bouldin Index (DBI) value. **Methods.** The clustering using the FCM method was conducted after first performing PCA on the data to address multicollinearity. Based on the PCA results, a total variance of 80% was used to represent the data's variability. Next, the data was clustered using FCM and evaluated with the DBI. **Results.** Based on the 80% total variance from the PCA results, four primary components were found to significantly represent the data's variability. From these four principal components, two clusters and three clusters were formed using the FCM method. Based on the DBI values, it was found that the two-cluster grouping was the best, with a DBI value of 0,088, compared to the three-cluster grouping, which had a DBI value of 1,226. The FCM clustering without PCA, excluding Makassar City from the data analysis, resulted in the best grouping of two clusters with a DBI value of 0.340. **Conclusion.** The FCM clustering with and without PCA on crime data showed that the best cluster, based on the DBI value, was obtained with two clusters, with a value of 0,088.

Keywords: *Principal Component Analysis, Davies Bouldin's Index, Fuzzy C-Means, Multicollinearity.*

DAFTAR LAMBANG/SINGKATAN

Lambang/Singkatan	Arti dan Penjelasan
FCM	<i>Fuzzy C-Means</i>
AKU	Analisis Komponen Utama
DBI	<i>Davies Bouldin's Index</i>
Y_i	Komponen utama ke- i
KMO	<i>Keiser Meyer Olkin</i>
χ^2	Uji <i>bartlett</i>
ρ	Matriks korelasi
U	Matriks keanggotaan
V	Matriks <i>centroid</i> data
P_t	Fungsi objektif ke- t
μ_{ik}	Bilangan keanggotaan data ke- i dan cluster ke- k
c	Banyak <i>cluster</i>
m	<i>fuzzifier</i>
<i>MaxIter</i>	Iterasi maksimum
$d_{ik}^2(x_i, v_k)$	Jarak antara vektor pengamatan ke- i dengan pusat kelompok ke- k
SSW_k	Jumlah kuadrat dalam <i>cluster</i> ke- k
$SSB_{k,c}$	Jumlah kuadrat antara <i>cluster</i> ke- k dan <i>cluster</i> ke- c

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGANTAR.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN.....	iv
UCAPAN TERIMA KASIH	v
PERNYATAAN KEASLIAN SKRIPSI.....	iv
DAN PELIMPAHAN HAK CIPTA.....	iv
ABSTRAK	vi
<i>ABSTRACT</i>	vii
DAFTAR LAMBANG/SINGKATAN	viii
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar belakang	1
1.2 Batasan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	3
1.5 Teori	3
1.5.1 Analisis <i>Cluster</i>	3
1.5.2 Analisis Komponen Utama	3
1.5.3 <i>Fuzzy C-Means</i>	7
1.5.4 <i>Fuzzy C-Means</i> dengan Analisis Komponen Utama	9
1.5.5 <i>Davies Bouldin's Index</i>	10
BAB II METODE PENELITIAN.....	14
2.1 Sumber Data.....	14
2.2 Struktur Data.....	14
2.3 Metode Penelitian	15
BAB III HASIL DAN PEMBAHASAN	17
3.1 Analisis Statistika Deskriptif	17

3.2 Analisis Komponen Utama	18
3.3 <i>Fuzzy C-Means</i>	22
3.3.1 <i>Fuzzy C-Means</i> pada Empat Komponen Utama	22
3.3.2 <i>Fuzzy C-Means</i> Tanpa Kota Makassar	31
3.3 Karakteristik Pengelompokan	41
3.4 Penentuan <i>Cluster</i> Terbaik	45
BAB IV KESIMPULAN	48
4.1 Kesimpulan	48
4.2 Saran	48
DAFTAR PUSTAKA	49
LAMPIRAN	52

DAFTAR TABEL

Tabel	Halaman
1. Tingkat Korelasi berdasarkan Nilai Koefisien Korelasi	5
2. Kriteria Nilai KMO.....	6
3. Variabel Penelitian.....	14
4. Struktur Data	15
5. Statistik Deskriptif Kriminalitas Kabupaten/Kota di Sulawesi Selatan.....	17
6. Uji KMO dan Uji <i>Bartlett</i>	19
7. Nilai Eigen, Proporsi Varians, dan Proporsi Kumulatif.....	20
8. Vektor Eigen.....	21
9. Skor pada Keempat Komponen Utama.....	22
10. Hasil Dua <i>Cluster</i> FCM Empat Komponen Utama Pertama	26
11. Hasil Tiga <i>Cluster</i> FCM Empat Komponen Utama Pertama.....	30
12. Uji KMO dan Uji <i>Bartlett</i> Tanpa Makassar	31
13. Hasil Dua <i>Cluster</i> FCM Tanpa Kota Makassar	36
14. Hasil Tiga <i>Cluster</i> FCM Tanpa Kota Makassar	40
15. Karakteristik Dua <i>Cluster</i> pada FCM dengan Empat Komponen Utama	41
16. Karakteristik Tiga <i>Cluster</i> pada FCM dengan Empat Komponen Utama.....	42
17. Karakteristik Dua <i>Cluster</i> pada FCM Tanpa Kota Makassar	43
18. Karakteristik Tiga <i>Cluster</i> pada FCM Tanpa Kota Makassar	44
19. Nilai DBI Setiap Pengelompokan	47

DAFTAR GAMBAR

Gambar	Halaman
1. Plot Uji Korelasi <i>Pearson</i>	19
2. Plot FCM Dua <i>Cluster</i> pada Empat Komponen Utama	26
3. Plot FCM Dua <i>Cluster</i> Tanpa Kota Makassar	31
4. Plot FCM Dua <i>Cluster</i> Tanpa Kota Makassar	36
5. Plot FCM Tiga <i>Cluster</i> Tanpa Kota Makassar.....	41

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Jumlah Kriminalitas 24 Kabupaten/Kota di Sulawesi Selatan.....	52
2. Data Standarisasi Kriminalitas 24 Kabupaten/Kota di Sulawesi Selatan.....	53
3. <i>Output</i> Hasil Uji Korelasi	54
4. <i>Output</i> Uji KMO dan Uji Batrlett.....	55
5. <i>Output</i> Hasil Analisis Komponen Utama.....	56
6. Skor Keempat Komponen Utama Pertama	58
7. Matriks keanggotaan awal U_0	59
8. Pusat <i>Cluster</i> Dua Cluster Empat Komponen Utama.....	61
9. Fungsi Objektif Pertama Fuzzy C-Means.....	63
10. Perubahan Matriks Keanggotaan U	65
11. <i>Output</i> Dua <i>Cluster Fuzzy C-Means</i> pada Empat Komponen Utama.....	67
12. <i>Output</i> Dua <i>Cluster Fuzzy C-Means</i> Tanpa Kota Makassar	68
13. <i>Output</i> Tiga <i>Cluster Fuzzy C-Means</i> pada Empat Komponen Utama	70
14. <i>Output</i> Tiga <i>Cluster Fuzzy C-Means</i> Tanpa Kota Makassar	71

BAB I

PENDAHULUAN

1.1 Latar belakang

Analisis *cluster* merupakan metode untuk mengelompokkan objek-objek data ke dalam suatu kelompok atau *cluster* sehingga objek yang berada dalam satu kelompok memiliki karakteristik yang mirip dibandingkan dengan objek pada *cluster* lainnya. Pengelompokan tersebut dilakukan atas dasar kemiripan atau jarak. Dalam kata lain, analisis *cluster* adalah metode analisis data yang bertujuan untuk membagi individu atau objek ke dalam beberapa kelompok yang memiliki karakteristik yang berbeda sehingga individu atau objek yang tergabung dalam satu kelompok yang sama akan memiliki sifat yang hampir homogen. Terdapat dua jenis pengelompokan yang paling umum yaitu secara hierarki dan non-hierarki (Johnson & Wichern, 2007).

Pengelompokan secara hierarki adalah suatu metode dalam analisis *cluster* yang secara bertahap mengelompokkan data atau objek berdasarkan kesamaan karakteristiknya sehingga terbentuk struktur berjenjang atau hierarki. Dalam proses ini, data dikelompokkan secara berurutan, mulai dari penggabungan *cluster* kecil hingga membentuk *cluster* yang lebih besar ataupun sebaliknya. Beberapa jenis analisis *cluster* secara hierarki terdiri dari *single linkage*, *complete linkage*, *average linkage*, dan metode *Ward*. Adapun pengelompokan secara non-hierarki digunakan untuk mengelompokkan objek dengan jumlah kelompok yang telah ditentukan sebelumnya. Dalam analisis *cluster* non-hierarki, metode yang paling umum digunakan adalah *K-Means* (Syafiyah dkk., 2023). Seiring berjalannya waktu, metode pengelompokan semakin berkembang, dan yang dikenal dewasa ini adalah metode pengelompokan *Fuzzy C-Means* dengan berbagai pengembangannya.

Metode pengelompokan *Fuzzy C-Means* (FCM) bertujuan untuk mengelompokkan setiap titik data berdasarkan kemiripan yang dimiliki oleh sebuah titik terhadap masing-masing *cluster* dengan suatu fungsi yang nilainya berada di antara nol dan satu. Setiap titik data akan memiliki nilai keanggotaan di setiap *cluster*, di mana nilai keanggotaan yang mendekati satu menunjukkan tingkat kemiripan yang tinggi antara sampel dengan kelompoknya, sedangkan nilai keanggotaan yang mendekati nol menunjukkan tingkat kemiripan yang rendah. Dengan demikian, FCM adalah metode pengelompokan data di dalam suatu *cluster* yang ditentukan oleh nilai derajat keanggotaan (Bezdek dkk., 1984).

Masalah yang biasa ditemukan dalam berbagai metode analisis *cluster* adalah jika terjadi korelasi yang kuat atau multikolinearitas antar data yang dapat mempengaruhi dalam analisis data. Hal ini juga disebabkan jika semakin banyak variabel yang terdapat dalam suatu gugus data, sehingga analisis data akan semakin kompleks dan informasi penting yang terdapat dalam data akan semakin sulit untuk diperoleh. Untuk itu maka Analisis Komponen Utama (AKU) dapat digunakan sebagai salah satu alternatif analisis untuk mengatasi multikolinearitas dan juga mereduksi dimensi yang ada pada data. Jumlah komponen utama lebih kecil dari jumlah variabel data awal. Selain itu, AKU dapat meningkatkan performa dari pengelompokan *fuzzy*.

Secara teknis, AKU dapat mereduksi dimensi pada suatu data sehingga dapat lebih efisien saat menjalankan algoritma pengelompokan *fuzzy*. Meskipun demikian, kualitas hasil pengelompokan masih dapat dipertahankan (Eliyanto & Sugiyarto, 2020).

Beberapa penelitian yang terdahulu mengenai analisis *cluster* telah dilakukan. Sanusi *dkk.* (2020) berdasarkan penelitiannya menggunakan FCM dalam mengelompokkan kabupaten/kota di Provinsi Sulawesi Selatan mendapatkan dua *cluster* terbaik untuk mengelompokkan data tersebut berdasarkan faktor penyebab gizi buruk. Sementara itu, Chen & Li (2019) dengan penelitiannya menyimpulkan bahwa pengelompokan *fuzzy* dapat dikombinasikan dengan AKU untuk meningkatkan efisiensi pengelompokan. Selain itu, untuk melihat seberapa baik pengelompokan yang dihasilkan dapat digunakan uji validitas *cluster* seperti yang digunakan oleh Amrulloh *dkk.* (2020) dalam penelitiannya menggunakan untuk mengevaluasi hasil cluster yang terbentuk diperoleh bahwa nilai *Davies Bouldin's Index* (DBI) menghasilkan akurasi pengelompokan yang lebih baik dibandingkan *Silhouette Coefficient* (SC).

Berdasarkan latar belakang tersebut maka peneliti mengambil judul "Penentuan *Cluster* Terbaik Menggunakan Metode *Fuzzy C-Means* dengan Analisis Komponen Utama", dan melihat hasil pengelompokan yang diperoleh. Dimensi data akan direduksi dengan menggabungkan variabel-variabel berkorelasi tinggi sehingga dapat meningkatkan efisiensi analisis dengan mempertahankan kualitas yang diperoleh dari hasil pengelompokan pada metode *Fuzzy C-Means*.

1.2 Batasan Masalah

Penelitian ini memiliki batasan masalah diantaranya diuraikan sebagai berikut:

1. Jenis data yang digunakan adalah data kriminalitas untuk 24 Kabupaten/Kota di Provinsi Sulawesi Selatan tahun 2022 yang berasal dari publikasi Badan Pusat Statistik Provinsi Sulawesi Selatan "Statistik Politik dan Keamanan Provinsi Sulawesi Selatan 2022".
2. Pada penelitian pendahuluan, pengelompokan dicoba untuk dua sampai lima cluster namun pengelompokan yang diperoleh tidak menggambarkan hasil yang diharapkan, sehingga pada penelitian ini dibatasi pengelompokan pada dua dan tiga *cluster*, dengan menggunakan nilai DBI sebagai ukuran untuk mengevaluasi kinerja pengelompokan.
3. Jumlah komponen utama dari AKU yang digunakan adalah yang menghasilkan total keragaman sebesar 80%.

1.3 Tujuan Penelitian

Berdasarkan permasalahan yang telah dijelaskan sebelumnya maka tujuan penelitian pada masalah yang akan dibahas pada penelitian ini diantaranya diuraikan sebagai berikut:

1. Mengelompokkan kabupaten/kota di Provinsi Sulawesi Selatan tahun 2022 berdasarkan data kriminalitas menggunakan metode *Fuzzy C-Means* dengan Analisis Komponen Utama.
2. Menentukan hasil pengelompokan terbaik berdasarkan metode *Fuzzy C-Means* dengan Analisis Komponen Utama berdasarkan nilai *Davies Bouldin's Index* (DBI).

1.4 Manfaat Penelitian

Dengan adanya penelitian yang dilakukan ini diharapkan dapat memberikan manfaat untuk berbagai pihak diantaranya menambah wawasan keilmuan terkait dengan analisis *cluster* sekaligus memberikan inspirasi dan gagasan baru bagi pembaca dalam mengembangkan analisis *cluster*. Selain itu, penelitian ini juga diharapkan dapat memberikan informasi mengenai pengelompokan 24 kabupaten/kota di Provinsi Sulawesi Selatan tahun 2022 berdasarkan jumlah kriminalitas menurut jenis kriminalitas sehingga dapat memberikan kontribusi dalam menangani kasus kriminalitas.

1.5 Teori

1.5.1 Analisis Cluster

Analisis *cluster* adalah jenis klasifikasi yang dikenakan pada sekumpulan objek yang terbatas (Jain & Dubes, 2000). Tujuan dari pengelompokan ini untuk menemukan kelompok-kelompok pengamatan yang terkait erat sehingga objek yang termasuk dalam *cluster* yang sama lebih mirip satu sama lain dibandingkan dengan objek yang termasuk dalam *cluster* lain. Dengan kata lain, analisis *cluster* mengelompokkan objek data hanya berdasarkan informasi yang ditemukan dalam data yang menggambarkan objek dan hubungannya (Tan dkk., 2006).

1.5.2 Analisis Komponen Utama

Analisis komponen utama pertama kali dikembangkan oleh Karl Pearson pada tahun 1901 (Hasan & Abdulazeez, 2021). Kemudian, dikembangkan lagi oleh Harold Hotelling pada tahun 1933 (Enzellina & Suhaedi, 2022). PCA merupakan teknik yang digunakan untuk mereduksi variabel berdimensi tinggi menjadi berdimensi rendah namun masih mengandung informasi dari variabel asal (Hasan & Abdulazeez, 2021). PCA dapat menggabungkan variabel-variabel yang berkorelasi tinggi dengan tujuan mengatasi adanya multikolinearitas (Daoud, 2018).

Menurut Johnson & Wichern, (2007), komponen utama adalah kombinasi linear dari p variabel acak X_1, X_2, \dots, X_p Dimana terdapat koefisien yang dinamakan sebagai vektor ciri atau vektor eigen. Vektor eigen dihasilkan dari akar ciri atau *eigen value* dari matriks kovarians (Σ) ataupun matriks korelasi (ρ) sehingga membuat komponen utama bergantung pada salah satu dari kedua matriks tersebut (Yulianto & Putriana, 2019).

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \quad (1)$$

dan

$$\rho = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix} \quad (2)$$

dimana Σ adalah matriks kovarians dengan σ_{ij} ; $i, j = 1, 2, \dots, p$ dan ρ adalah matriks korelasi dengan ρ_{ij} ; $i, j = 1, 2, \dots, p$.

Misalkan terdapat vektor acak $X' = [X_1, X_2, \dots, X_p]$ memiliki Σ dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ sehingga dapat terbentuk kombinasi linear dengan

$$\begin{aligned} Y_1 &= \mathbf{a}'_1 X = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ Y_2 &= \mathbf{a}'_2 X = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ &\vdots \\ Y_p &= \mathbf{a}'_p X = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned}$$

Komponen utama adalah kombinasi linear Y_1, Y_2, \dots, Y_p yang tidak berkorelasi. Misalkan Σ memiliki nilai eigen-vektor eigen berpasangan $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ dimana $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ maka komponen utama ke- i dapat ditulis dengan

$$Y_i = \mathbf{e}'_i X = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p \quad (3)$$

dimana

$$\begin{aligned} \text{Var}(Y_i) &= \mathbf{e}'_i \Sigma \mathbf{e}_i = \lambda_i \\ \text{Cov}(Y_i, Y_k) &= \mathbf{e}'_i \Sigma \mathbf{e}_k = 0 \end{aligned}$$

dimana $i = 1, 2, \dots, p$; $i \neq k$; Jika beberapa λ_i adalah sama maka pemilihan vektor sesuai adalah \mathbf{e}_i dan akibatnya Y_i tidak unik. Berdasarkan hal tersebut, komponen utama menjadi tidak berkorelasi dan mempunyai varians yang sama dengan nilai eigen dari Σ (Johnson & Wichern, 2007). Adapun keragaman total komponen utama berdasarkan populasi dapat dituliskan:

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad (4)$$

Keragaman total yang mampu dijelaskan oleh setiap komponen utama adalah proporsi antara akar ciri komponen utama tersebut dengan jumlah akar ciri

matriks kovarians (Yulianto & Putriana, 2019). Proporsi keragaman total yang disebabkan komponen utama ke- i adalah:

$$\text{Total varian komponen utama ke } - i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (5)$$

Jumlah komponen utama yang diambil adalah sebanyak q buah dimana $q < p$ (Yulianto & Putriana, 2019). Pada umumnya banyak komponen utama bergantung batas ketetapan total varian komponen utama yang dapat dihasilkan. Secara umum, proporsi keragaman yang biasa digunakan adalah minimal 80%, beberapa juga menggunakan hingga 90%. Komponen-komponen utama yang dipilih akan dapat menggantikan variabel p yang asli tanpa menghilangkan informasi dari variabel asal (Johnson & Wichern, 2007).

1.5.2.1 Uji Korelasi Pearson

Uji korelasi *pearson* diperlukan untuk melihat seberapa besar hubungan antar variabel. Nilai korelasi yang tinggi mengindikasikan adanya multikolinearitas antar variabel. Pendeteksian adanya korelasi antar variabel dilakukan dengan menghitung koefisien korelasi sederhana atau korelasi *pearson* antar variabel bebas (Azman dkk., 2021). Penentuan koefisien korelasi dapat digunakan dengan persamaan:

$$R_{ij} = \frac{n \sum_{i,j=1}^n X_i X_j - \sum_{i=1}^n X_i \sum_{j=1}^n X_j}{\sqrt{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \sqrt{n \sum_{j=1}^n X_j^2 - (\sum_{j=1}^n X_j)^2}} \quad (6)$$

Keterangan:

- R_{ij} : Koefisien korelasi antara variabel ke- i dan ke- j
- n : Banyak data
- X_i : Variabel ke- i
- X_j : Variabel ke- j

Koefisien korelasi (R_{ij}) berisikan nilai dalam interval antara + 1 dan - 1, termasuk nilai akhir + 1 atau - 1. Banyak sumber yang mengkategorikan kuat lemahnya koefisien korelasi. Menurut Ratner (2009), nilai dari koefisien korelasi dapat dijelaskan sebagai berikut.

Tabel 1. Tingkat Korelasi berdasarkan nilai Koefisien Korelasi.

Nilai Koefisien Korelasi	Tingkat Korelasi
$R_{ij} = 0$	Tidak Ada
$0 < R_{ij} \leq 0.3$	Lemah
$0.3 < R_{ij} \leq 0.7$	Sedang
$0.7 < R_{ij} < 1$	Kuat
$ R_{ij} = 1$	Sempurna

Tabel 1 menunjukkan terjadinya korelasi yang kuat antar variabel ketika nilai koefisien korelasi sebesar $0.7 < |R_{ij}| < 1$ dan $|R_{ij}| = 1$ jika kedua variabel berkorelasi sempurna.

1.5.2.2 Uji Kaiser Meyer Olkin dan Uji Bartlett

Uji *Kaiser Meyer Olkin* atau uji KMO digunakan untuk menguji kelayakan dari data observasi yang dimiliki sebelum dilakukan AKU (Noya van Delsen dkk., 2017). Dengan kata lain, uji ini digunakan untuk menguji kecukupan data yang digunakan untuk difaktorkan (Purnomo dkk., 2022). Menurut Shrestha (2021), uji KMO diberikan dalam persamaan:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p R_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p R_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (7)$$

dengan $i = 1, 2, \dots, p; j = 1, 2, \dots, p;$

dimana:

R_{ij} : Koefisien korelasi sederhana antara variabel ke- i dan ke- j

a_{ij} : Koefisien korelasi parsial antara variabel ke- i dan ke- j

Nilai a_{ij} dapat diperoleh dari dekomposisi matriks invers dari R dan matriks Diagonal dari R^{-1} (Matriks A disebut sebagai Matriks Anti-Image Korelasi) dengan persamaan

$$A = R^{-1}DR^{-1} \quad (8)$$

dengan R^{-1} adalah invers dari matriks korelasi R^{-1} dan D matriks Diagonal dari R^{-1} .

Nilai KMO bervariasi mulai dari 0 hingga 1. Nilai KMO akan mendekati 1 apabila koefisien korelasi lebih besar daripada koefisien korelasi parsial (Nitasari dkk., 2023). Menurut Kaiser (1960), interpretasi untuk kriteria uji KMO dijabarkan pada tabel berikut.

Tabel 2. Kriteria Nilai KMO

Nilai KMO	Interpretasi
0,90 – 1,00	Data sangat baik
0,80 – 0,90	Data baik
0,70 – 0,80	Data agak cukup
0,60 – 0,70	Data lebih dari cukup
0,50 – 0,60	Data cukup
0,00 – 0,50	Data tidak layak

Selain itu, perlu juga dilakukan uji *Bartlett* yang bertujuan untuk melihat hubungan antar variabel dalam kasus multivariat. Jika variabel X_1, X_2, \dots, X_p bersifat independen atau saling bebas maka matriks korelasi antar variabel sama dengan matriks

identitas (Purnomo dkk., 2022). Menurut Shrestha (2021), Uji *Bartlett* dinyatakan dengan hipotesis sebagai berikut:

Hipotesis:

$H_0 : \mathbf{R} = \mathbf{I}$ (matriks korelasi merupakan matriks identitas)

$H_1 : \mathbf{R} \neq \mathbf{I}$ (matriks korelasi bukan merupakan matriks identitas)

Statistik Uji:

$$\chi^2 = - \left(n - 1 - \left(\frac{2p + 5}{6} \right) \right) \ln(\det(\mathbf{R})) \quad (9)$$

dimana:

p : Jumlah variabel

n : Jumlah observasi

\mathbf{R} : Matriks korelasi

Keputusan menerima H_0 jika $\chi_{hitung}^2 < \chi_{\sigma.p(p-1)/2}^2$, dan menolak H_0 jika $\chi_{hitung}^2 \geq \chi_{\sigma.p(p-1)/2}^2$. Matriks korelasi dari variabel-variabel yang diuji jika merupakan matriks identitas maka tidak ada korelasi linear antara variabel-variabel tersebut.

1.5.3 Fuzzy C-Means

Fuzzy C-Means (FCM) pertama kali ditemukan oleh Dunn pada tahun 1973 dan dikembangkan lagi oleh Bezdek pada tahun 1981 (Nurrachman, 2021). FCM menggunakan konsep dari bidang logika *fuzzy* dan teori himpunan *fuzzy* dalam melakukan pengelompokan (Tan dkk., 2006), dimana dalam pengelompokan *fuzzy* memungkinkan data untuk menjadi anggota dari semua kelas atau *cluster* dengan tingkat keanggotaan yang berbeda-beda, biasanya dalam rentang antara 0 hingga 1. Hal itu menunjukkan bahwa setiap data dapat memiliki tingkat keanggotaan yang berbeda terhadap setiap kelompok (Suryadi, 2015).

Konsep pada FCM pada awalnya menentukan pusat *cluster* yang akan menjadi penanda lokasi untuk rata-rata setiap *cluster* dimana pusat *cluster* ini mungkin tidak akurat pada awalnya dan seringkali dilakukan secara acak. Setelah itu, setiap data diberikan *membership function* atau derajat keanggotaan (u_{ik}) untuk setiap *cluster* yang memiliki jangkauan nilai $0 \leq u_{ik} \leq 1$. Derajat keanggotaan ini menentukan seberapa dekat suatu data dengan masing-masing *cluster*. Sebagian besar pengguna FCM memilih *fuzzifier* atau nilai m dalam rentang $(1, 10]$ dimana $m = 2$ sejauh ini merupakan pilihan yang paling umum. Terakhir adalah menentukan fungsi objektif dan melakukan perulangan sehingga lokasi *cluster* bisa berada dalam posisi yang stabil (Bezdek, 1995; Suryadi, 2015). FCM merupakan algoritma yang efektif, namun nilai acak pada titik pusat dapat menyebabkan proses iteratif cenderung menuju solusi lokal yang optimal dengan mudah. Pemilihan dan penyesuaian ukuran ketidaksetaraan memegang peranan penting dalam struktur

cluster data (Surono & Putri, 2021). Secara rincinya, algoritma FCM dapat dilakukan dengan beberapa tahap sebagai berikut (Sanusi dkk., 2020):

1. Mengorganisir data yang akan di-*cluster* \mathbf{X} dalam bentuk matriks $n \times m$ dimana n adalah jumlah data dan m adalah jumlah atribut setiap data. Matriks X_{ij} mewakili nilai data ke- k ($i = 1, 2, \dots, n$) untuk atribut ke- j ($j = 1, 2, \dots, m$).
2. Menentukan banyak *cluster* (c), *fuzzifier* (m), iterasi maksimum (*MaxIter*), perubahan nilai fungsi objektif terkecil yang diharapkan (ε), fungsi objektif awal ($P_0 = 0$), dan iterasi awal ($t = 1$).
3. Menentukan bilangan random μ_{ik} yang mewakili data ke- i ($i = 1, 2, \dots, n$) dan *cluster* ke- k ($k = 1, 2, \dots, c$) yang dijadikan elemen-elemen awal pada matriks keanggotaan awal \mathbf{U} .

$$\mathbf{U}_0 = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nc} \end{pmatrix} \quad (10)$$

Dimana harus memenuhi beberapa kondisi diantaranya:

- a. $\mu_{ik} = [0, 1]$; $1 \leq i \leq c; 1 \leq k \leq n$
 - b. $\sum_{i=1}^n \mu_{ik} = 1$; $1 \leq i \leq c$
 - c. $0 < \sum_{i=1}^c \mu_{ik} < c$; $1 \leq k \leq n$
4. Menghitung pusat *cluster* atau *centroid* ke- k dengan persamaan:

$$V_{ij} = \frac{\sum_{i=1}^n ((\mu_{ik})^m X_{kj})}{\sum_{i=1}^n (\mu_{ik})^m} \quad (11)$$

dimana bentuk matriks yang akan dihasilkan:

$$\mathbf{V} = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{c1} & \cdots & v_{cm} \end{bmatrix}$$

5. Menentukan fungsi objektif pada iterasi ke- t , P_t dengan persamaan:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^m d_{ik}^2(x_i, v_k) \quad (12)$$

dengan $d_{ik}^2(x_i, v_k)$ adalah jarak antara vektor pengamatan ke- i dengan pusat kelompok ke- k .

6. Menghitung perubahan dari nilai matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^p (Y_{ij} - V_{kj})^2]^{-\frac{1}{m-1}}}{\sum_{k=1}^c [\sum_{j=1}^p (Y_{ij} - V_{kj})^2]^{-\frac{1}{m-1}}} \quad (13)$$

7. Mengecek kondisi berhenti dengan persamaan:

$$|P_t - (P_t - 1)| < \varepsilon \quad (14)$$

atau $t > \text{MaxIter}$, dimana *MaxIter* adalah maximum iterasi yang digunakan.

Jika tidak $t = t + 1$ maka kembali pada langkah ke-4.

1.5.3.1 Jarak *Euclidean Kuadrat*

Metode pengelompokan pada dasarnya menggunakan konsep jarak antar objek pengamatan. Jarak antara data dan *centroid* dapat diukur dengan menggunakan

metrik jarak. Beberapa metode untuk menghitung jarak antar titik yang pada umumnya digunakan untuk algoritma seperti pengelompokan, segmentasi, dan sebagainya. Salah satu dari jarak metrik yang biasa digunakan adalah Jarak *Euclidean* (Sathasivam & Sagir, 2020). Jarak *Euclidean* tidak berubah ketika nilai yang sama ditambahkan ke setiap variabel data, namun berubah jika dikalikan. Dengan kata lain, penggunaan fungsi kuadrat (d^2) pada $d = (x_k - y_i)$ berpengaruh pada perbedaan untuk nilai yang lebih kecil dari 1 atau yang lebih besar dari 1. Oleh karena itu, jarak *euclidean* dapat diubah menjadi jarak *euclidean* kuadrat yang dihitung dengan tidak menggunakan akar kuadrat pada jarak *Euclidean*. Menurut Ullsch & Lötsch (2022), jarak tersebut disebut Jarak Euclidean Kuadrat (*Square-Euclidean Distances*) dengan persamaan

$$d(x_i, y_k) = \sqrt{\sum_{i=1}^n (x_i - v_k)^2}$$

$$d^2(x_i, v_k) = \sum_{i=1}^n (x_i - v_k)^2 \quad (15)$$

Keterangan:

$d_{ik}^2(x_i, v_k)$: jarak antara vektor pengamatan ke- i dengan pusat kelompok ke- k .

x_i : Vektor pengamatan ke- i

v_k : Pusat kelompok ke- k .

1.5.4 Fuzzy C-Means dengan Analisis Komponen Utama

Pengelompokan *Fuzzy C-Means* (FCM) dapat dikombinasikan dengan AKU. Hampir sama dengan algoritma FCM pada umumnya, tetapi pengelompokan memakai hasil dari komponen utama berisi sebagian besar informasi dari variabel asli dan berisi informasi yang lebih padat. AKU menggabungkan variabel-variabel yang saling berkorelasi kuat menjadi komponen utama yang bersifat independen secara linier. Setiap komponen utama yang dihasilkan akan memaksimalkan keragaman dalam data. Dengan cara ini, AKU memungkinkan representasi data dalam dimensi yang lebih rendah dengan tetap mempertahankan sebanyak mungkin informasi keragaman dari data asli. Pengelompokan FCM merupakan metode pengelompokan sampel data secara *fuzzy*. Derajat keanggotaan setiap objek ke pusat pengelompokan diperoleh dengan mengoptimalkan fungsi objektif, yang memungkinkan sampel menjadi anggota kelompok yang berbeda dengan derajat keanggotaan tertentu dengan sifat-sifat yang serupa dapat dikelompokkan ke dalam satu kelompok. Pengelompokan FCM yang dikombinasikan dengan AKU, selain mendapatkan evaluasi sampel akhir, tetapi juga untuk mengelompokkan sampel berdasarkan kemiripannya (Chen & Li, 2019). Dalam beberapa studi kasus, performa FCM dapat meningkat setelah diterapkan AKU pada data. Dengan dimensi yang lebih rendah, pengelompokan FCM akan menjadi lebih efisien sekaligus

menghilangkan masalah dari korelasi yang kuat antar variabel (Eliyanto & Sugiyarto, 2020).

1.5.5 Davies Bouldin's Index

Davies Bouldin's Index (DBI) adalah salah satu metode validasi dalam analisis *cluster*. DBI mengukur rasio antara jarak antar *cluster* dengan jarak dalam *cluster*. Penggunaan DBI dapat digunakan untuk memvalidasi pengelompokan data pada setiap *cluster* (Mustofa & Iman Saufik Suasana, 2020). Jika jarak antar *cluster* maksimum (bernilai besar), kemiripan karakteristik antara masing-masing *cluster* cenderung kecil, sehingga perbedaan antar *cluster* lebih jelas. Jika jarak intra-*cluster* minimum, maka setiap data dalam *cluster* tersebut memiliki tingkat kemiripan karakteristik yang tinggi. Suatu *cluster* dikatakan memiliki pengelompokan yang baik jika memiliki nilai DBI yang minimum atau mendekati 0 (Suroño & Putri, 2021). Nilai DBI dapat dihitung menggunakan persamaan:

$$DBI = \frac{1}{k} \sum_{c=1}^k \max_{k \neq c} Ratio_{k,c} \quad (16)$$

Dimana

$$Ratio_{k,c} = \frac{SSW_k + SSW_c}{SSB_{k,c}}$$

$$SSW_k = \frac{1}{m_k} \sum_{j=1}^{m_k} d(x_j, v_k)$$

$$SSB_{k,c} = d(v_k, v_c)$$

Keterangan:

m_k : Jumlah data dalam *cluster* ke- k

x : Data pada *cluster*

$v_{k,c}$: *Centroid cluster* ke- k dan ke- c

SSW_k : Jumlah kuadrat dalam *cluster* ke- k

$SSB_{k,c}$: Jumlah kuadrat antara *cluster* ke- k dan *cluster* ke- c

$Ratio_{k,c}$: Ukuran rasio nilai perbandingan *cluster* ke- k dan *cluster* ke- c

DBI : Nilai *Davies Bouldin's Index*

1.5.6 Kriminalitas

Kriminalitas atau tindak kejahatan adalah fenomena sosial yang secara terus-menerus dihadapi oleh masyarakat. Ada berbagai faktor yang menyebabkan seseorang melakukan tindak kejahatan, termasuk faktor ekonomi, lingkungan, keluarga, dan faktor kepribadian individu. Kriminalitas sering kali dianggap sebagai

hasil dari pembangunan yang tidak merata, disproportional, dan terjadi baik di wilayah perkotaan maupun pedesaan (Utami & Asih, 2021).

Upaya pencegahan kejahatan pada hakikatnya adalah bagian dari upaya perlindungan masyarakat. Oleh karena itu, tujuan pokok kebijakan pidana adalah perlindungan yang ditujukan untuk menjamin kesejahteraan masyarakat. Kebijakan atau upaya yang ditujukan untuk menjamin sosial merupakan bagian integral dari kebijakan sosial (Silaen & Siregar, 2020).

Kejahatan atau pelanggaran adalah tindakan yang dapat dikenakan hukuman berdasarkan Ketentuan Hukum Pidana atau Undang-Undang serta peraturan lainnya yang berlaku di Indonesia. Pembukaan Undang-Undang Dasar 1945, pada alinea keempatnya, secara eksplisit menyatakan tujuan negara Indonesia, salah satunya adalah "Melindungi segenap bangsa Indonesia dan seluruh tumpah darah Indonesia". Dengan demikian, negara Indonesia berkomitmen untuk melindungi semua elemen di Indonesia, termasuk rakyat, kekayaan alam, kekayaan budaya, dan nilai-nilai luhur bangsa. Sejalan dengan tujuan tersebut, negara memiliki tanggung jawab untuk menjamin keselamatan dan keamanan warganya dari segala bentuk kejahatan sebagai upaya untuk mencapai tujuan perlindungan bangsa. Dalam pendekatan hukum, kejahatan diartikan sebagai tindakan yang melanggar hukum pidana atau Undang-Undang yang berlaku dalam masyarakat. Di Indonesia, definisi kejahatan dalam lembaga penegak hukum didasarkan pada Ketentuan Undang-Undang Hukum Pidana (KUHP) atau Undang-Undang lain yang relevan. (Badan Pusat Statistik, 2023).

Menurut Badan Pusat Statistik (2023), klasifikasi jenis kejahatan atau tinfak kriminalitas dapat disajikan dengan mengikuti pedoman dari Kitab Undang-Undang Hukum Pidana (KUHP) Republik Indonesia dan *International Classification of Crime for Statistical Purposes (ICCS)* yang berasal dari gagasan *United Nations Office on Drugs and Crime (UNODC)*. Berikut ini adalah klasifikasi jenis kriminalitas yang disajikan berdasarkan publikasi Badan Pusat Statistik:

1. Kejahatan terhadap nyawa
Hak untuk hidup adalah salah satu hak asasi manusia yang paling penting, menurut *Universal Declaration of Human Rights*. Pembunuhan adalah tindakan untuk menghilangkan nyawa seseorang merupakan kejahatan yang tergolong berat hukumnya.
2. Kejahatan terhadap Fisik/Badan
Penganiayaan berat, penganiayaan ringan, dan kekerasan dalam rumah tangga adalah contoh kejahatan fisik/badan. Penganiayaan adalah perbuatan dengan sengaja merusak kesehatan fisik orang lain baik menimbulkan penyakit luka, cacat, sakit atau halangan untuk menjalankan pekerjaan sehari-hari maupun tidak.
3. Kejahatan terhadap kesusilaan
Klasifikasi kejahatan terhadap kesusilaan terdiri dari jenis kejahatan perkosaan dan pencabulan. Perkosaan adalah tindakan pemaksaan hubungan seksual tanpa persetujuan korban sedangkan pencabulan adalah

perbuatan yang menyerang kehormatan kesusilaan dengan melakukan sentuhan di bagian tubuh tertentu yang dilakukan secara sepihak dan tidak diharapkan oleh korban.

4. **Kejahatan terhadap kemerdekaan orang**
Klasifikasi kejahatan terhadap kemerdekaan orang terdiri dari jenis kejahatan penculikan dan mempekerjakan anak dibawah umur. Penculikan merupakan tindakan mencuri atau melarikan seseorang lalu disembunyikan dan biasanya dimintakan tebusan.
5. **Kejahatan terhadap hak milik/barang dengan penggunaan kekerasan**
Dalam hierarki klasifikasi jenis kejahatan, jenis kejahatan terhadap hak milik/barang dengan penggunaan kekerasan tergolong tinggi karena selain menyangkut kejahatan terhadap properti, juga merupakan kejahatan terhadap fisik. Klasifikasi kejahatan terhadap hak milik/barang dengan penggunaan kekerasan terdiri dari pencurian dengan kekerasan, pencurian dengan kekerasan menggunakan senjata api (senpi), dan pencurian dengan kekerasan menggunakan senjata tajam (sajam). Pencurian dengan kekerasan adalah mengambil sesuatu barang atau ternak yang disertai kekerasan atau ancaman kekerasan.
6. **Kejahatan terhadap hak milik/barang tanpa Penggunaan Kekerasan**
Kejahatan terhadap hak milik/barang tanpa penggunaan kekerasan terdiri dari pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, pengrusakan/penghancuran barang, pembakaran dengan sengaja, dan penadahan.
7. **Kejahatan terkait narkotika**
Kejahatan terkait narkotika merupakan jenis kejahatan yang termasuk kejahatan luar biasa. Kelompok kejahatan terkait narkotika mencakup kejahatan narkotika dan psikotropika. Tindakannya dapat berupa menawarkan untuk dijual, menjual, membeli, menerima, menjadi perantara dalam jual beli, menukar, menyerahkan, atau menerima narkotika dan psikotropika.
8. **Kejahatan terkait penipuan, penggelapan, dan korupsi**
Klasifikasi kejahatan ini mencakup kejahatan penipuan, penggelapan, dan korupsi. Pasal 378 KUHP mendefinisikan penipuan sebagai menguntungkan diri sendiri atau orang lain secara melawan hukum dengan menggunakan nama palsu, martabat palsu dengan tipu muslihat. Perbedaan korupsi dan penggelapan terletak pada kepemilikan obyek tindak pidana tersebut. Harta atau benda yang diambil dalam tindak pidana penggelapan adalah harta milik orang lain dan bukan negara, sedangkan harta atau benda yang diambil dalam tindak pidana korupsi adalah harta atau uang negara.
9. **Kejahatan terhadap ketertiban umum**
Klasifikasi ini segala sesuatu konflik dapat diartikan sebagai interaksi antara dua pihak atau lebih, baik individu maupun kelompok, yang memiliki atau merasa memiliki tujuan yang tidak sejalan. konflik sebagai suatu proses sosial di mana individu atau kelompok berupaya mencapai tujuan mereka dengan

cara menantang pihak lawan, seringkali disertai ancaman dan/atau kekerasan sehingga dapat mengganggu ketertiban umum.

BAB II METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian adalah jumlah kasus kriminalitas untuk 24 kabupaten/kota di Provinsi Sulawesi Selatan tahun 2022 yang berasal dari publikasi Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan “Statistik Politik dan Keamanan Provinsi Sulawesi Selatan 2022” yang dapat diakses di situs resmi BPS Provinsi Sulawesi Selatan yaitu <https://sulsel.bps.go.id>. Data tersebut terdiri dari jenis-jenis kriminalitas atau kejahatan dari 24 kabupaten/kota di Provinsi Sulawesi Selatan pada tahun 2022 dilampirkan pada Lampiran 1. Adapun variabel-variabel yang digunakan dalam penelitian:

Tabel 3. Variabel Penelitian

Variabel	Keterangan
X_1	Pembunuhan
X_2	Penganiayaan Berat (Anirat)
X_3	Penganiayaan Ringan (Anira)
X_4	Kekerasan Dalam Rumah Tangga (KDRT)
X_5	Perkosaan
X_6	Melanggarkesopanan/kesusilaan/pencabulan
X_7	Penculikan
X_8	Pencurian dengan Kekerasan (Curas)
X_9	Pencurian biasa
X_{10}	Pencurian Kendaraan Bermotor (Curanmor)
X_{11}	Pencurian dengan Pemberatan (Curat)
X_{12}	Pengrusakan/Penghancuran Barang
X_{13}	Pembakaran dengan Sengaja
X_{14}	Narkotika dan Psikotropika
X_{15}	Penipuan/Perbuatan Curang
X_{16}	Penggelapan
X_{17}	Korupsi
X_{18}	Kejahatan terhadap Ketertiban Umum

2.2 Struktur Data

Variabel-variabel yang digunakan dalam penelitian ini yaitu jenis-jenis kriminalitas dari 24 kabupaten/kota di Provinsi Sulawesi Selatan. Gambaran struktur data yang digunakan dijabarkan pada Tabel 4.

Tabel 4. Struktur Data

Kabupaten/kota	Variabel (j)					
(k)	X_1	X_2	...	X_j	...	X_{18}
1	$X_{1,1}$	$X_{1,2}$...	$X_{1,j}$...	$X_{1,18}$
2	$X_{2,1}$	$X_{2,2}$...	$X_{2,j}$...	$X_{2,18}$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
k	$X_{k,1}$	$X_{k,2}$...	$X_{k,j}$...	$X_{k,18}$
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
24	$X_{24,1}$	$X_{24,2}$...	$X_{24,j}$...	$X_{24,18}$

2.3 Metode Penelitian

Langkah-langkah analisis data yang dilakukan berdasarkan tujuan penelitian sebagai berikut:

1. Memperoleh hasil data baru berdasarkan variansi kumulatif Analisis Komponen Utama (AKU).
 - a. Melakukan uji korelasi *Pearson* untuk mengecek terjadinya korelasi kuat antar variabel.
 - b. Melakukan Uji *Kaiser Meyer Olkin* (KMO) dan Uji *Bartlett* untuk melihat apakah data cukup layak digunakan untuk AKU atau tidak.
 - c. Melakukan AKU dan mengambil komponen utama berdasarkan nilai total keragaman kumulatif sebesar 80%.
 - d. Mendapatkan nilai skor dari vektor eigen berdasarkan variansi kumulatif yang dihasilkan sebagai data baru.
2. Menentukan hasil pengelompokan terbaik dengan metode *Fuzzy C-Means* (FCM).
 - a. Mengorganisir data baru dengan variansi kumulatif 80% yang akan di-*cluster* yaitu Y dalam bentuk matriks $n \times m$ dimana n adalah jumlah data dan m adalah jumlah atribut setiap data. Matriks Y_{kj} mewakili nilai data ke- k ($k = 1, 2, \dots, n$) untuk atribut ke- j . ($j = 1, 2, \dots, m$)
 - b. Menentukan banyak *cluster* (c) dimana dalam penelitian ini dibatasi dengan dua dan tiga *cluster*, *fuzzifier* (m), iterasi maksimum (*MaxIter*), perubahan nilai fungsi objektif terkecil yang diharapkan (ϵ), fungsi objektif awal ($P_0 = 0$), dan iterasi awal ($t = 1$)
 - c. Menentukan bilangan random μ_{ik} dimana $i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$ yang dijadikan elemen-elemen awal pada matriks keanggotaan awal U .

$$U_0 = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{24,1} & \mu_{24,2} & \cdots & \mu_{24,c} \end{pmatrix}$$

Dimana harus memenuhi beberapa kondisi diantaranya:

$$\begin{aligned} \mu_{ik} &= [0,1] && ; 1 \leq i \leq c; 1 \leq k \leq n \\ \sum_{i=1}^n \mu_{ik} &= 1 && ; 1 \leq i \leq c \\ 0 < \sum_{i=1}^c \mu_{ik} &< c && ; 1 \leq k \leq n \end{aligned}$$

- d. Menghitung pusat *cluster* atau *centroids* ke- k dengan persamaan:

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^m Y_{kj})}{\sum_{i=1}^n (\mu_{ik})^m}$$

dimana bentuk matriks yang akan dihasilkan:

$$V = \begin{bmatrix} v_{11} & \dots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{c1} & \dots & v_{cm} \end{bmatrix}$$

- e. Menentukan fungsi objektif pada iterasi ke- t , P_t dengan persamaan:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^m d_{ik}^2(y_k, v_i)$$

dengan $d_{ik}^2(x_k, v_i)$ adalah jarak antara vektor pengamatan ke- k dengan pusat kelompok ke- i yaitu jarak *Euclidian* kuadrat

$$d^2(y_k, v_i) = \sum_{i=1}^n (y_k - v_i)^2$$

- f. Menghitung perubahan dari nilai matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^p (Y_{ij} - V_{kj})^2]^{-\frac{1}{m-1}}}{\sum_{k=1}^c [\sum_{j=1}^p (Y_{ij} - V_{kj})^2]^{-\frac{1}{m-1}}}$$

- g. Mengecek kondisi berhenti dengan persamaan:

$$|P_t - (P_t - 1)| < \varepsilon$$

atau $t >$ Iterasi maksimum. Jika tidak $t = t + 1$ maka kembali pada langkah d.

- h. Melakukan kembali langkah pertama kembali jika terdapat kabupaten/kota yang membentuk kelompok sendiri. Namun, jika tidak memenuhi Uji KMO atau Uji *Bartlett* maka langsung dikelompokkan dengan FCM.

- i. Menentukan hasil pengelompokan terbaik dari *cluster* optimum yang dihasilkan dengan menggunakan nilai *Davies Bouldin's Index*.