

**SKRIPSI**  
**ANALISIS KARAKTERISTIK *CLUSTER* BERDASARKAN**  
**FAKTOR-FAKTOR KEMISKINAN DI PROVINSI SULAWESI**  
**SELATAN MENGGUNAKAN *K-MEDOIDS CLUSTERING***



**SARANTY**  
**E0221508**

**PROGRAM STUDI STATISTIKA**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS SULAWESI BARAT**  
**2025**

## SURAT PERNYATAAN

Yang bertanda tangan dibawah ini

Nama : Saranty

Tempat/Tgl. Lahir : Mamasa, 15 September 2001

NIM : E0221508

Program Studi : Statistika

Menyatakan bahwa tugas akhir dengan judul “Analisis Karakteristik *Cluster* berdasarkan Faktor-Faktor Kemiskinan di Provinsi Sulawesi Selatan menggunakan *Clustering K-Medoids*” disusun berdasarkan prosedur ilmiah yang telah melalui pembimbingan dan bukan merupakan plagiat dari karya ilmiah/naskah yang lain. Apabila dikemudian hari terbukti bahwa pernyataan ini tidak benar, maka bersedia menerima sanksi sesuai peraturan yang berlaku.

Majene, 23 September 2025



HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Saranty

Nim : E0221508

Judul : Analisis karakteristik *cluster* berdasarkan faktor-faktor kemiskinan di Provinsi Sulawesi Selatan menggunakan *K-medoids Clustering*

Telah berhasil dipertanggungjawabkan dihadapan Tim Penguji (SK Nomor 89/UN55.7/HK.04/2023, tanggal 23 Juli 2025) dan diterima sebagai bagian persyaratan memperoleh gelar sarjana S1 pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sulawesi Barat.

Disahkan Oleh:

Dekan FMIPA  
Universitas Sulawesi Barat

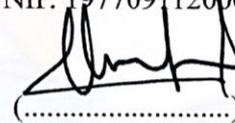


Musafira, S.Si., M.Sc.

NIP. 19770911200660422002

Tim penguji:

Ketua penguji : Musafira, S.Si., M.Sc.



(.....)

Sekretaris : Muh. Hijrah, S.Pd., M.Si.



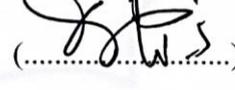
(.....)

Pembimbing 1 : Muhammad Hidayatullah, S.Pd., M.Kom



(.....)

Pembimbing 2 : Fardinah, S.Si., M.Sc



(.....)

Penguji 1 : Putri Indi Rahayu, S.Si., M.Stat.



(.....)

Penguji 2 : Retno Mayapada, S.Si., M.Si



(.....)

Penguji 3 : Darma Ekawati, S.Si., M.Si.



(.....)

## ABSTRAK

Kemiskinan merupakan masalah sosial yang kompleks dan masih menjadi tantangan utama di berbagai daerah di Indonesia, termasuk Provinsi Sulawesi Selatan. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Sulawesi Selatan berdasarkan sembilan indikator kemiskinan, yaitu indeks kedalaman kemiskinan, indeks keparahan kemiskinan, tingkat pengangguran terbuka, umur harapan hidup, Indeks Pembangunan Manusia (IPM), pengeluaran per kapita, rata-rata lama sekolah, Produk Domestik Regional Bruto (PDRB), dan kepadatan penduduk. Metode yang digunakan adalah *K-Medoids Clustering* dengan pendekatan *Data Mining* dan *Machine Learning*. Hasil evaluasi menggunakan metrik internal seperti *Silhouette Coefficient* dan *Dunn Index* menunjukkan bahwa jumlah *cluster* optimal adalah dua ( $K=2$ ), dengan kualitas cluster yang sangat baik (*Silhouette Coefficient* sebesar 0.6369, *Dunn Index* 1.0227). Analisis karakteristik menunjukkan bahwa *cluster* pertama berisi kabupaten/kota dengan indikator kemiskinan yang lebih tinggi, sedangkan *cluster* kedua mencakup wilayah dengan karakteristik pembangunan yang lebih maju. Selain itu, analisis tren dari tahun 2019 hingga 2023 menunjukkan bahwa sebagian besar kabupaten/kota memiliki komposisi *cluster* yang stabil, namun terjadi perubahan signifikan pada tahun 2023 dengan berpindahnya kota Parepare dan kota Palopo dari *cluster* 1 ke *cluster* 2. Temuan ini diharapkan dapat menjadi dasar dalam merancang kebijakan pengentasan kemiskinan yang lebih tepat sasaran di Sulawesi Selatan.

**Kata kunci:** Kemiskinan, *K-Medoids*, *Clustering*, Sulawesi Selatan, Evaluasi *Cluster*.

## **ABSTRACT**

*Poverty is a complex social problem and is still a major challenge in various regions in Indonesia, including South Sulawesi Province. This study aims to group districts/cities in South Sulawesi based on nine poverty indicators, namely poverty depth index, poverty severity index, open unemployment rate, life expectancy, human development index (HDI), per capita expenditure, average length of schooling, gross regional domestic product (GRDP), and population density. The method used is K-Medoids Clustering with a data mining and machine learning approach. The evaluation results using internal metrics such as Silhouette Coefficient and Dunn Index show that the optimal number of clusters is two ( $K = 2$ ), with very good cluster quality (Silhouette Coefficient of 0.6369 Dunn Index 1.0227). Characteristic analysis shows that the first cluster contains districts/cities with higher poverty indicators, while the second cluster includes areas with more advanced development characteristics. In addition, trend analysis from 2019 to 2023 shows that most districts/cities have a stable cluster composition, but there will be a significant change in 2023 with the movement of Parepare and Palopo cities from cluster 1 to cluster 2. It is hoped that these findings can be the basis for designing more targeted poverty alleviation policies in South Sulawesi.*

*Keywords: Poverty, K-Medoids, Clustering, South Sulawesi, Cluster Evaluation.*

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kemiskinan merupakan suatu keadaan dimana ketidakmampuan seseorang dalam mencukupi kebutuhan primer seperti sandang, pangan, papan, pendidikan, maupun kesehatan. Kemiskinan adalah salah satu masalah sosial yang dihadapi oleh berbagai negara di dunia, termasuk Indonesia. Isu ini menjadi tantangan global yang tidak membedakan antara negara maju dan negara berkembang (Hidayat dkk., 2022).

Masalah tentang kemiskinan merupakan masalah lama yang belum terselesaikan, meskipun pemerintah sudah begitu banyak memberikan bantuan bagi penduduk kurang mampu, seperti Jaminan Kesehatan Nasional, Kartu Indonesia Sehat (KIS), Kartu Indonesia Pintar (KIP) serta jenis bantuan lainnya. Namun hal itu masih belum mampu mengatasi sepenuhnya masalah kemiskinan di Indonesia. Oleh karena itu, pengentasan masalah kemiskinan harus menjadi prioritas utama dalam pembangunan ekonomi, baik jangka pendek maupun jangka panjang (Hamdiah dkk., 2024).

Tingkat kemiskinan di Indonesia menunjukkan tren fluktuatif dalam lima tahun terakhir. Pada tahun 2018, jumlah penduduk miskin tercatat sebesar 9,98 juta jiwa dan menurun menjadi 9,41 juta jiwa pada 2019. Namun, pada tahun 2020 terjadi peningkatan menjadi 9,78 juta jiwa, yang kemudian melonjak cukup signifikan pada 2021 menjadi 10,14 juta jiwa, kemungkinan besar disebabkan oleh dampak ekonomi pandemi COVID-19. Selanjutnya, angka kemiskinan mulai menurun menjadi 9,54 juta jiwa pada 2022, dan kembali turun menjadi 9,36 juta jiwa pada tahun 2023 (Badan Pusat Statistik, 2023). Berdasarkan data tersebut, tingkat kemiskinan di Indonesia masih tergolong tinggi dan memerlukan perhatian serius.

Salah satu Provinsi di Indonesia yang memiliki tingkat kemiskinan tinggi adalah Provinsi Sulawesi Selatan. Berdasarkan hasil Survei Sosial Ekonomi Nasional (Susenas) Maret 2024, jumlah penduduk miskin di Sulawesi Selatan

berjumlah 736,48 ribu jiwa atau 0,06 persen dari total penduduk miskin (BPS, 2024). Adapun faktor-faktor yang digunakan Badan Pusat Statistik (BPS) dalam menentukan tingkat kemiskinan adalah indeks kedalaman kemiskinan, indeks keparahan kemiskinan, tingkat pengangguran terbuka, umur harapan hidup, Indeks Pembangunan Manusia (IPM), pengeluaran per kapita, rata rata lama sekolah, Produk Domestik Regional Bruto (PDRB), dan kepadatan penduduk. Setiap kabupaten memiliki kombinasi nilai yang berbeda-beda untuk faktor-faktor tersebut, sehingga karakteristik kemiskinan antarwilayah menjadi kompleks dan multidimensional. Untuk memahami pola kemiskinan dan mengidentifikasi wilayah-wilayah yang memiliki karakteristik serupa, diperlukan pendekatan statistik. Salah satu metode yang dapat digunakan adalah pengelompokan (*clustering*).

*Clustering* merupakan metode analisis data yang sering digunakan untuk mengelompokkan data dalam kelompok-kelompok tertentu berdasarkan karakteristik yang ditentukan. Tujuan dari metode ini adalah untuk mengelompokkan data dengan karakteristik serupa ke dalam satu kelompok, sementara data dengan karakteristik yang berbeda ditempatkan ke kelompok yang berbeda. Algoritma *K-Medoids* merupakan bagian dari *clustering* yang sama dengan algoritma *K-Means*. Kelebihan dari metode ini mampu mengatasi kelemahan dari metode *K-Means* yang sensitif terhadap *outlier*. Adapun perbedaan algoritma *K-Medoids* dan *K-Means* terdapat pada pemilihan data point sebagai pusatnya (Arbain dkk., 2023).

Penelitian yang dilakukan oleh Sinatrya (2018) membandingkan algoritma *K-Medoids* dan *K-Means*. Hasil analisis keseluruhan dalam penelitian ini menunjukkan akurasi rata-rata algoritma *K-Medoids* mempunyai tingkat akurasi lebih besar dibandingkan dengan *K-Means*. Kemudian Safitri dkk (2021) penelitiannya bertujuan untuk mengetahui kelompok miskin menurut kabupaten/kota dengan menggunakan metode *Clustering K-Medoids* yang mampu mengatasi data yang sensitif terhadap *outlier*, penelitian ini menghasilkan tiga kelompok, di mana kelompok pertama terdiri dari delapan wilayah, kelompok kedua mencakup 15 wilayah, dan kelompok ketiga memiliki empat wilayah.

Selanjutnya, penelitian yang dilakukan oleh Senang Hati Gulo (2023) mengelompokkan penduduk di Kecamatan Afulu menggunakan metode *K-Medoids* dan menghasilkan tiga *cluster* yaitu mampu, cukup, dan kurang mampu. Hasil evaluasi DBI sebesar 0,1502 menunjukkan *cluster* cukup baik. Lebih lanjut Zahra dkk (2024) dengan pengelompokan tingkat kemiskinan di setiap Provinsi di Indonesia menggunakan Algoritma *K-Medoids*, menghasilkan 3 *cluster* dari Nilai *Silhouette Score* yaitu 0,4 yang menunjukkan hasil dari *clusternya* sangat baik. Dari keempat penelitian tersebut, maka dapat disimpulkan bahwa Algoritma *K-Medoids* merupakan algoritma yang cukup fleksibel untuk diterapkan diberbagai tujuan pengelompokan garis kemiskinan. Namun, yang menjadi pembeda peneliti terdahulu dengan penelitian ini adalah penelitian terdahulu tidak menganalisis pola perubahan komposisi *cluster* dari tahun ke tahun.

Berdasarkan permasalahan diatas, penulis mengangkat judul “Analisis Karakteristik *Cluster* berdasarkan faktor-faktor kemiskinan di Provinsi Sulawesi Selatan menggunakan *K-Medoids Clustering*”. Harapan dari penelitian ini dapat memberikan wawasan yang lebih komprehensif mengenai kemiskinan, tidak hanya dalam bentuk pengelompokan wilayah, tetapi juga dengan memahami karakteristik setiap *cluster* berdasarkan faktor-faktor yang mempengaruhi tingkat kemiskinan.

## 1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang di atas, maka rumusan masalah untuk penelitian ini yaitu

1. Bagaimana pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor kemiskinan menggunakan *K-Medoids Clustering*?
2. Bagaimana karakteristik antar *cluster* berdasarkan hasil analisis profil *cluster*?
3. Bagaimana pola perubahan komposisi *cluster* kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor kemiskinan dari tahun 2019 hingga 2023?

## 1.3 Tujuan Penelitian

Penelitian ini memiliki beberapa tujuan yaitu :

1. Untuk mengetahui hasil pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor kemiskinan menggunakan metode *K-Medoids Clustering*
2. Untuk mengetahui karakteristik *cluster* yang terbentuk dari hasil pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor kemiskinan menggunakan metode *K-Medoids Clustering*
3. Untuk mengetahui pola perubahan komposisi *cluster* kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan faktor-faktor kemiskinan dari tahun 2019 hingga 2023

#### **1.4 Manfaat Penelitian**

Penelitiann ini memiliki manfaat penelitian yaitu :

1. Penelitian ini akan membantu mengidentifikasi faktor-faktor utama yang memengaruhi tingkat kemiskinan di Sulawesi Selatan. Pemahaman yang lebih mendalam tentang faktor-faktor ini dapat membantu berbagai pihak dalam merumuskan kebijakan atau program yang lebih efektif .
2. Secara teoritis, penelitian ini diharapkan memberikan dampak yang positif terhadap algoritma *K-Medoids* dalam penerapannya di berbagai tujuan *cluster*.
3. Secara metodologis, penelitian ini diharapkan menjadi rujukan dalam mengelompokan penduduk dengan menggunakan algoritma *K-Medoids*.

#### **1.5 Batasan Masalah**

Ada beberapa batasan masalah dalam peneltian ini, yaitu :

1. Objek yang akan diteliti adalah 24 kabupaten/kota di Provinsi Sulawesi Selatan.
2. Penerapan analisis *cluster* yang akan digunakan dalam penelitian ini adalah dengan menggunakan analisis *K-Medoids Clustering*.
3. Penelitian ini menggunakan 9 variabel faktor-faktor kemiskinan. Kesembilan variabel tersebut yaitu Indeks kedalaman kemiskinan, indeks keparahan kemiskinan, tingkat pengangguran terbuka, umur harapan hidup, Indek Pembangunan Manusia (IPM), pengeluaran per kapita, rata-

rata lama sekolah, Produk Domestik Regional Bruto (PDRB), dan kepadatan penduduk .

## BAB II TINJAUAN PUSTAKA

### 2.1 Kemiskinan

Kemiskinan tetap menjadi salah satu persoalan sosial yang kompleks di Indonesia hingga saat ini (Muntahanah dkk., 2024). Berdasarkan sifatnya, kemiskinan terbagi menjadi 2 jenis yaitu kemiskinan absolut dan kemiskinan relatif. Kemiskinan absolut mengacu pada kondisi masyarakat yang penghasilannya dibawah batas minimum untuk memenuhi kebutuhan dasar seperti pangan, sandangan dan papan. Sementara itu, kemiskinan relatif menggambarkan kondisi dimana meskipun masyarakat sudah dapat memenuhi kebutuhan dasar minimum, penghasilan mereka masih rendah di bandingkan dengan standar hidup masyarakat di sekitarnya (Adji dkk., 2020).

Kemiskinan, secara umum terjadi saat pendapatan seseorang tidak cukup untuk mencukupi kebutuhan hidup. Dalam publikasi Badan Pusat Statistik (BPS) menggunakan pendekatan *basic needs* yang juga diterapkan di berbagai negara. Dengan pendekatan ini, kemiskinan diartikan sebagai ketidakmampuan secara ekonomi untuk mencukupi kebutuhan dasar, baik makanan maupun non-makanan, yang diukur berdasarkan pengeluaran (BPS, 2023).

Penduduk yang tergolong miskin adalah mereka yang memiliki rata-rata pengeluaran per kapita per bulan di bawah Garis Kemiskinan (GK) yang dihitung dari hasil survei berbasis sampel. Angka kemiskinan yang dipublikasikan oleh BPS bersifat makro dan didasarkan pada data dari Survei Sosial Ekonomi Nasional (Susenas). Data ini menggambarkan persentase penduduk miskin terhadap total populasi di suatu wilayah (Widianingsih dkk., 2022).

*World Bank* (2001) dalam publikasinya, *World Development Report 2000/2001: Attacking Poverty*, menjelaskan bahwa kemiskinan adalah kondisi di mana standar hidup yang layak tidak terpenuhi. Selain itu *World Bank* menentukan kemiskinan berdasarkan kurangnya sandang, pangan, papan, layanan kesehatan, dan akses pendidikan.

## 2.2 Faktor-Faktor Kemiskinan

Dalam penelitian ini, Faktor-faktor utama yang memengaruhi kemiskinan dapat diidentifikasi melalui sejumlah variabel yang diperoleh dari data Badan Pusat Statistik (BPS), antara lain:

- a) Indeks kedalaman kemiskinan, adalah ukuran yang menunjukkan seberapa jauh rata-rata pengeluaran penduduk miskin terhadap garis kemiskinan. Semakin dalam tingkat kemiskinan yang dialami, semakin besar kebutuhan intervensi untuk mengangkat penduduk keluar dari kemiskinan. Ini mencerminkan tingkat kerentanan ekonomi yang lebih tinggi.
- b) Indeks keparahan kemiskinan, mengukur ketimpangan di antara penduduk miskin, ketimpangan di antara penduduk miskin menunjukkan adanya kelompok sangat miskin yang kondisinya lebih parah, dan ini membuat program pengentasan kemiskinan menjadi lebih kompleks dan sulit ditangani secara merata.
- c) Tingkat pengangguran terbuka, adalah persentase angkatan kerja yang tidak memiliki pekerjaan tetapi aktif mencari kerja. Tanpa pekerjaan, individu tidak memiliki sumber pendapatan, yang merupakan penyebab langsung kemiskinan. Tingginya pengangguran juga menandakan rendahnya akses ke pasar kerja yang produktif.
- d) Umur harapan hidup, rata-rata usia seseorang berdasarkan kondisi kesehatan dan pelayanan medis di suatu daerah. Semakin tinggi angkanya, semakin baik kualitas hidup penduduk. Umur harapan hidup yang rendah menunjukkan lemahnya layanan kesehatan, gizi, dan lingkungan yang mendukung kehidupan
- e) Indeks Pembangunan Manusia (IPM) ukuran yang menggambarkan kualitas pembangunan manusia, IPM yang rendah mencerminkan keterbatasan kapasitas manusia untuk berpartisipasi dalam pembangunan ekonomi, sehingga meningkatkan risiko kemiskinan yang berkelanjutan.
- f) Pengeluaran per kapita, rata-rata jumlah pengeluaran individu dalam suatu rumah tangga dalam periode tertentu, sebagai indikator kesejahteraan

ekonomi. Pengeluaran rendah menunjukkan keterbatasan dalam memenuhi kebutuhan dasar, yang merupakan ciri utama kemiskinan.

- g) Rata-rata lama sekolah, adalah jumlah rata-rata tahun yang telah dihabiskan oleh penduduk berusia 25 tahun ke atas dalam mengikuti pendidikan formal. Pendidikan yang rendah membatasi peluang kerja dan pendapatan yang layak, serta kemampuan untuk keluar dari siklus kemiskinan.
- h) Produk Domestik Regional Bruto (PDRB), merupakan total nilai barang dan jasa yang dihasilkan di suatu daerah dalam periode tertentu. Wilayah dengan PDRB rendah biasanya memiliki aktivitas ekonomi terbatas, sehingga pendapatan masyarakat juga rendah dan risiko kemiskinan meningkat.
- i) Kepadatan Penduduk, tingginya kepadatan penduduk bisa menyebabkan tekanan pada sumber daya alam, infrastruktur, dan layanan publik. Dalam kondisi ini, distribusi sumber daya menjadi tidak optimal dan kualitas hidup menurun (Badan Pusat Statistik, 2023)

Kesembilan variabel ini digunakan untuk mengelompokkan wilayah berdasarkan karakteristik kemiskinannya melalui metode *Clustering K-Medoids*, dengan tujuan memahami pola kemiskinan yang beragam di Sulawesi Selatan.

### **2.3 Data Mining**

*Data mining* adalah proses eksplorasi atau penemuan informasi baru dengan cara mengidentifikasi pola atau aturan tertentu dari kumpulan data besar yang terakumulasi. Proses ini juga dapat diartikan sebagai langkah-langkah sistematis untuk menggali nilai tambah dari data berupa pengetahuan tersembunyi yang sebelumnya tidak dapat ditemukan secara manual, namun memiliki manfaat yang signifikan (Nazeriandy dkk., 2021)

*Data mining* juga dapat diartikan sebagai teknik untuk menemukan informasi dan pola-pola bermanfaat dari kumpulan data dalam jumlah besar. Tahapan dalam *data mining* meliputi pengumpulan, ekstraksi, analisis, serta pengolahan data secara statistik. Istilah ini juga sering disebut sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan berbagai sebutan lainnya (Arhami dan Nasir, 2020).

Dari definisi tersebut, *data mining* dapat disimpulkan sebagai teknik untuk mengekstrak informasi berharga yang tersembunyi dalam kumpulan data (*database*) berukuran besar, dengan tujuan menemukan pola-pola menarik yang sebelumnya tidak diketahui. Kata "*mining*" sendiri menunjukkan proses memperoleh nilai dari sejumlah besar bahan dasar.

Dalam penelitian ini, *data mining* digunakan untuk menganalisis faktor-faktor yang berkontribusi terhadap tingkat kemiskinan di wilayah tertentu. Teknik ini membantu dalam mengidentifikasi pola-pola yang relevan sehingga dapat menjadi dasar untuk pengelompokan wilayah berdasarkan karakteristik kemiskinan. Dalam penelitian ini, data mining menjadi kerangka besar di mana metode *clustering* (termasuk *K-Medoids*) digunakan. Dengan menggunakan data *mining*, penelitian ini dapat mengidentifikasi pola kemiskinan, membandingkan daerah yang memiliki karakteristik serupa, hal ini bertujuan untuk mempermudah perencanaan dan pelaksanaan program pengentasan kemiskinan yang lebih tepat sasaran.

#### **2.4 Machine Learning**

Menurut Hindrayani dkk., (2021), *Machine Learning* adalah studi tentang algoritma yang dirancang untuk mempelajari dan melaksanakan berbagai hal tertentu yang secara otomatis dilakukan oleh manusia. Proses belajar di sini mengacu pada kemampuan untuk menyelesaikan berbagai tugas atau membuat prediksi yang akurat berdasarkan pola-pola yang telah dipelajari sebelumnya. Sedangkan menurut Hariyono (2020) semakin sering *machine learning* digunakan, tingkat akurasi akan semakin meningkat seiring berjalannya waktu. Hal ini disebabkan oleh proses pembelajaran yang terus berkembang melalui interaksi dengan pengguna.

*Machine Learning* adalah algoritma yang berfungsi sebagai sarana interaksi antara mesin dan pengguna (*user*) dalam menganalisis serta memodelkan data untuk menemukan solusi suatu permasalahan. *machine learning* terbagi menjadi 2 (dua) pendekatan utama yaitu *supervised learning* adalah metode yang menggunakan data yang sudah memiliki label atau output yang diketahui. Saat diuji dengan data baru, model ini menghasilkan output sesuai dengan pola yang

dipelajari dari data latihan. Sedangkan *unsupervised learning* adalah metode pembelajaran tanpa data berlabel. Model berusaha menemukan pola atau struktur dalam data secara mandiri. Teknik ini mirip dengan cara kerja otak manusia, di mana pembelajaran dilakukan tanpa contoh eksplisit, sehingga model harus mengenali pola tersembunyi untuk menyelesaikan masalah dengan data baru (Fahmi, 2023).

Dalam konteks *clustering* faktor kemiskinan di Sulawesi Selatan, *machine learning* memungkinkan pengelompokan wilayah yang memiliki karakteristik yang serupa. Dengan demikian, *machine learning* tidak hanya membantu dalam pengolahan data tetapi juga memberikan wawasan yang dapat digunakan untuk perencanaan kebijakan yang lebih tepat sasaran.

*Machine Learning* pada penelitian ini berperan sebagai kerangka metodologis untuk mengolah data kemiskinan, menganalisis pola-pola yang terdapat dalam data, dan membagi data ke dalam beberapa kelompok berdasarkan karakteristiknya. Teknik *unsupervised learning* sangat sesuai untuk penelitian ini karena data yang digunakan tidak memiliki label tertentu, dan tujuan utama adalah untuk memahami struktur data kemiskinan. Teknik *clustering*, sebagai salah satu pendekatan *unsupervised learning* dalam *machine learning* digunakan untuk mengelompokkan data tanpa label yang telah ditentukan sebelumnya.

## **2.5 Metode Clustering**

*Clustering* merupakan salah satu teknik dalam *data mining* yang digunakan untuk mengelompokkan objek-objek ke dalam beberapa *cluster*. *Cluster* sendiri merujuk pada kelompok objek data yang memiliki kesamaan tinggi satu sama lain dalam *cluster* yang sama, dan perbedaan yang jelas dengan objek-objek di *cluster* lainnya. Proses ini bertujuan untuk mengelompokkan objek-objek ke dalam satu atau lebih *cluster*, dimana objek dalam *cluster* yang sama memiliki kemiripan yang signifikan. *Clustering* merupakan metode untuk mengidentifikasi dan mengelompokkan data yang memiliki kesamaan karakteristik (*similarity*) antara satu data dengan data lainnya (Martanto dkk.,2022).

*Clustering* adalah proses pengelompokan *record*, observasi, atau objek ke dalam kelompok tertentu. *Clustering* sering digunakan sebagai tahap awal dalam

*data mining* sebelum menerapkan metode analisis yang lebih lanjut. Terdapat berbagai algoritma *clustering* yang telah dikembangkan oleh peneliti, seperti *K-Means*, *Improved K-Means*, *Fuzzy C-Means*, *DBSCAN*, *K-Medoids (PAM)*, *CLARANS*, dan *Fuzzy Subtractive*. Masing-masing algoritma memiliki kelebihan dan kekurangannya, namun pada dasarnya bekerja dengan prinsip yang sama, yaitu mengelompokkan data berdasarkan karakteristik tertentu dan mengukur jarak atau kemiripan antar data dalam satu kelompok (Syukron dkk., 2022).

Pengelompokan data bertujuan untuk mengelompokkan objek yang serupa, menemukan pola dalam data, serta memperdalam analisis dan pemahaman data. Beberapa metode *clustering* yang umum digunakan termasuk *K-Means*, *K-Medoids*, *Hierarchical Clustering*, dan *Density-Based Clustering (DBSCAN)*. Dalam penelitian ini, metode *K-Medoids Clustering* dipilih sebagai pendekatan yang relevan untuk analisis data.

## **2.6 Metode K-Medoids**

*Algoritma Partitioning Around Medoids (PAM)*, yang juga dikenal sebagai algoritma *K-Medoids*, adalah algoritma yang menggunakan *medoid* sebagai wakil dari *cluster*. Perbedaan utama antara algoritma *K-Medoids* dan *K-Means* terletak pada pemilihan pusat *cluster*, *K-Medoids* menggunakan objek nyata sebagai *medoid* untuk masing-masing *cluster*, sementara *K-Means* menggunakan nilai rata-rata (mean) sebagai pusat *cluster* (Harahap, 2021).

*K-Medoids* adalah salah satu algoritma *clustering* yang bertujuan untuk membagi atau mengelompokkan data menjadi beberapa kelompok atau bagian. Berbeda dari teknik pembelajaran terawasi (*supervised learning*), di mana data diklasifikasikan berdasarkan nilai yang telah ditetapkan sebelumnya seperti dalam analisis sentimen, di mana data awal dikelompokkan sebagai sentimen positif atau negatif algoritma *K-Medoids* melakukan pengelompokan secara mandiri berdasarkan data yang ada, dan karena itu termasuk dalam kategori pembelajaran tanpa pengawasan (*unsupervised learning*). Input yang diperlukan adalah nilai  $k$ , yang menunjukkan jumlah *cluster* yang diinginkan, serta data yang akan dikelompokkan. *Metode Clustering* ini menggunakan *medoid* sebagai pusat *clusternya*. *Medoid* adalah objek yang berada di pusat sebuah *cluster*. Metode ini

memiliki keunggulan dalam hal ketahanan, yaitu tidak mudah terpengaruh oleh keberadaan data pencilan atau data ekstrim (Bhat dalam Fajrina, 2021).

Berikut adalah beberapa langkah dalam algoritma *K-Medoids* (Rahmah dkk., 2022)

1. Menentukan pusat *cluster* awal secara acak sebanyak  $k$  (jumlah *cluster*) sesuai jumlah *cluster* yang diinginkan.
2. Menghitung jarak setiap objek ke *cluster* terdekat menggunakan persamaan jarak *Euclidean Distance*.
3. Pemilihan kandidat *medoid* baru
4. Hitung jarak setiap objek di *cluster* terhadap kandidat non-*Medoid*.
5. Ulangi langkah 3-4 hingga tidak ada perubahan pada *medoid*.

## 2.7 Euclidean Distance

*Euclidean distance* adalah metode pengukuran jarak yang mengacu pada jarak antara dua titik di suatu dimensi (Zhu dkk., 2021). Penelitian ini menggunakan *Euclidean Distance* karena metrik ini adalah salah satu ukuran jarak yang paling sederhana dan intuitif untuk menghitung kedekatan antar data. *Euclidean Distance* mengukur jarak lurus antara dua titik dalam ruang multidimensi, sehingga mampu merepresentasikan perbedaan antar variabel dengan baik. Hal ini penting untuk metode *K-Medoids*, yang bergantung pada *Medoid* sebagai representasi pusat *cluster*, di mana jarak antar data menjadi dasar utama untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik. Selain itu, *Euclidean Distance* sederhana secara komputasi dan efektif dalam menangkap hubungan geometris antar data, menjadikannya pilihan yang ideal untuk penelitian ini.

Kita dapat menghitung jarak *Euclidean Distance* dengan persamaan (2.1)

$$d(i, j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2.1)$$

dengan :

$d(i, j)$  : Jarak *Euclidean* antara titik  $i$  dan  $j$ .

$x_{ik}$  dan  $x_{jk}$  : Koordinat ke- $k$  dari masing-masing titik data

$n$  : Jumlah dimensi atribut yang digunakan

## 2.8 Silhouette Coefficient

*Silhouette Coefficient* adalah metrik yang digunakan untuk menilai kualitas *clustering* dalam analisis data. Metrik ini mengevaluasi sejauh mana setiap objek data lebih sesuai dengan *clusternya* di bandingkan dengan *cluster* lain. Koefisien *silhouette* mengintegrasikan konsep kohesi (kedekatan dalam satu *cluster*) dan separasi (pemisahan antar *cluster*). Nilainya berkisar antara -1 hingga 1, di mana *cluster* dianggap baik jika nilai *Silhouette Coefficient* mendekati 1 (Atira dkk., 2023)

*Silhouette Coefficient* merupakan metode yang fungsinya untuk menilai hasil *cluster* dalam analisis data. *Silhouette Coefficient* mengevaluasi sejauh mana setiap objek data sesuai dengan *clusternya* sendiri dibandingkan dengan *cluster* lainnya. Langkah-langkah yang digunakan dalam metode *Silhouette Coefficient*

1. Menghitung rata-rata jarak data satu dengan data yang lainnya yang terdapat pada satu *cluster* yang sama menggunakan persamaan (2.2)

$$a_i = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.2)$$

dengan :

$a(i)$  : Perbedaan rata-rata objek (i) ke semua objek lain pada *cluster* A

$a(i, j)$  : Jarak antara data I dengan data j

A : *Cluster*

2. Menghitung rata-rata jarak dari data tersebut dengan semua data di *cluster* lain menggunakan persamaan (2.3)

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (2.3)$$

dengan :

$d(i, C)$  : Perbedaan rata-rata objek I ke semua objek lain pada C

C : *Cluster* lain selain *cluster* A atau *cluster* C tidak sama dengan *cluster*

A

3. Menghitung  $d(i,C)$  untuk semua  $C$ , maka diambil nilai terkecil menggunakan persamaan(2.4)

$$b(i) = \min_{c \neq A} d(i, C) \quad (2.4)$$

dengan :

$b(i)$  : *Cluster* tetangga dari objek  $i$  yang mencapai nilai minimum

*Cluster*  $B$  yang mencapai minimum yaitu ( $d(i, B) = b(i)$ ) di sebut tetangga dari objek ( $i$ ) dan ini adalah *cluster* terbaik kedua untuk objek ( $i$ )

4. Menghitung nilai *Silhouette Coefficient* dapat dilihat pada persamaaan (2.5)

$$s(i) = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (2.5)$$

dengan :

$s(i)$  : nilai *Silhouette Coefficient*

$a_i$  : jarak rata-rata  $i$  dengan seluruh objek pada *cluster* yang sama

$b_i$  : jarak rata-rata  $i$  dengan seluruh objek pada *cluster* lain

## 2.9 Dunn Index

*Dunn Index* adalah rasio jarak terkecil antara observasi pada kelompok yang berbeda dengan jarak terbesar pada masing-masing *cluster* data. *Dunn Index* diperoleh dari hasil pembagian antara  $d_{\min}$  dengan  $d_{\max}$ .

$$c = \frac{d_{\min}}{d_{\max}} \quad (2.6)$$

dengan:

$d_{\min}$  : jarak terkecil antara observasi pada kelompok yang berbeda

$d_{\max}$  : jarak terbesar pada masing-masing kelompok

*Dunn Index* bertujuan untuk mengidentifikasi *cluster* yang terpisah dengan baik (Qadrini, L. 2020).

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian mengenai analisis karakteristik *cluster* berdasarkan faktor-faktor kemiskinan di Provinsi Sulawesi Selatan menggunakan metode *K-Medoids Clustering*, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Hasil pengelompokan kabupaten/kota di Provinsi Sulawesi Selatan berhasil dikelompokkan ke dalam dua *cluster* utama berdasarkan sembilan variabel kemiskinan, yaitu indeks kedalaman kemiskinan, indeks keparahan kemiskinan, tingkat pengangguran terbuka, umur harapan hidup, Indeks Pembangunan Manusia (IPM), pengeluaran per kapita, rata-rata lama sekolah, Produk Domestik Regional Bruto (PDRB), dan kepadatan penduduk.
2. Karakteristik tiap *cluster* dari tahun 2019-2023 :
  - 1) *Cluster* 1 dikategorikan sebagai wilayah tertinggal, yang ditandai dengan rendahnya rata-rata IPM, umur harapan hidup, pengeluaran per kapita, rata-rata lama sekolah, dan PDRB, serta tingginya indeks kedalaman dan keparahan kemiskinan.
  - 2) *Cluster* 2 mencakup wilayah maju, yang memiliki capaian sosial ekonomi yang lebih tinggi secara signifikan pada semua indikator yang dianalisis, termasuk kepadatan penduduk yang jauh lebih tinggi.
3. Pola Perubahan Komposisi *cluster* (2019–2023): Selama periode 2019 hingga 2022, tidak terjadi perubahan komposisi *cluster* di antara kabupaten/kota, menunjukkan kondisi sosial ekonomi yang cenderung stagnan. Namun pada tahun 2023, terjadi perubahan signifikan, Kota Parepare, dan Kota Palopo berpindah dari *cluster* 1 ke *cluster* 2. Perubahan ini mengindikasikan adanya peningkatan kualitas pembangunan sosial-

ekonomi, seperti penurunan tingkat kemiskinan, peningkatan IPM, serta akses pendidikan dan kesehatan yang lebih baik.

Dengan demikian, hasil penelitian ini memberikan gambaran yang jelas mengenai pemetaan wilayah berdasarkan kemiripan karakteristik kemiskinan, yang dapat dijadikan dasar dalam perumusan kebijakan pembangunan daerah yang lebih tepat sasaran dan berkeadilan.

## **5.2 Saran**

Berdasarkan hasil penelitian, disarankan agar pemerintah daerah menggunakan hasil *cluster* ini sebagai dasar kebijakan pengentasan kemiskinan, khususnya di wilayah yang termasuk dalam *cluster* tertinggal. Perlu dilakukan pembaruan data secara berkala untuk memantau perubahan kondisi sosial ekonomi daerah. Penelitian selanjutnya disarankan untuk menggunakan metode *clustering* lain atau menambahkan variabel yang lebih beragam guna memperoleh hasil yang lebih komprehensif. Selain itu, dibutuhkan kolaborasi lintas sektor untuk merancang program yang sesuai dengan karakteristik setiap *cluster*.

### DAFTAR PUSTAKA

- Adji, A., Hidayat, T., Tuhiman, H., Kurniawati, S., & Maulana, A. (2020). Pengukuran Garis Kemiskinan di Indonesia: Tinjauan Teoretis dan Usulan Perbaikan. Jakarta: Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K).
- Arbain, D., Sriyanto, S., & Triloka, J. (2023, August). Perbandingan Kinerja Algoritma *K-Medoids* Dan *K-Means* Untuk Klasifikasi Penyakit Kanker Serviks. In *Prosiding Seminar Nasional Darmajaya* (Vol. 1, pp. 118-131).
- Arhami, M., & Nasir, M. (2020). *Data Mining: Algoritma dan Implementasinya dalam Dunia Nyata*. Yogyakarta: Penerbit Andi.
- Atira, A., & Sari, B. N. (2023). Penerapan *Silhouette Coefficient*, *Elbow Method* dan *Gap Statistics* untuk Penentuan Cluster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan. *Jurnal Ilmiah Wahana Pendidikan*, 9(17), 76-86.
- Badan Pusat Statistik, 2018-2023 <https://sulsel.bps.go.id/id>
- Badan Pusat Statistik. (2023). *Produk Domestik Regional Bruto Kabupaten/Kota di Indonesia 2018–2022*
- Fahmi, M. N. (2023). Implementasi *Machine Learning* menggunakan *Python Library: Scikit-Learn (Supervised dan Unsupervised Learning)*. *Sains Data Jurnal Studi Matematika dan Teknologi*, 1(2), 87-96.
- Fajrina, E. A. N. (2021). Analisis *Cluster Robust* Menggunakan Metode *K-Medoids* Pada Data Mengandung Pencilan.
- Hamdiah, V. (2024). Peran kebijakan fiskal dalam mengevaluasi zakat sebagai upaya menyikapi kemiskinan di Indonesia. *Jurnal Ilmiah Ekonomi Islam*, 10(1), 333-339.
- Han, J., Kamber, M., and Pei, J., 2011. *Data Mining Concepts and Techniques*. (3rd ed.). USA: Morgan Kaufmann.
- Harahap, F. (2021). Perbandingan Algoritma *K-Means* dan *K-Medoids* untuk *Clustering* Kelas Siswa Tunagrahita. *TIN: Terapan Informatika Nusantara*, 2(4), 193–194.
- Hariyono, F. P. P. (2020). Penerapan *Machine Learning* Untuk Prediksi Penyakit Jantung Menggunakan Metode Naïve Bayes Dan Decision Tree (Doctoral dissertation, UNIVERSITAS AIRLANGGA).
- Hidayat, F. P., Putra, R. P., Alfitrah, M. D., & Widodo, E. (2023). Implementasi *Clustering K-Medoids* dalam Pengelompokan Kabupaten di Provinsi Aceh Berdasarkan Faktor yang Mempengaruhi Kemiskinan. *Indonesian Journal of Applied Statistics*, 5(2), 121-130.
- Hindrayani, K. M., Anjani, A., & Nurlaili, A. L. (2021, October). Penerapan

*Machine Learning* pada Penjualan Produk UMKM: Studi Literatur. In *Prosiding Seminar Nasional Sains Data* (Vol. 1, No. 01, pp. 19-23).

- Kurniawan, R., Martadinata, A. T., & Cahyo, S. D. (2023). Klasifikasi Tingkat Kematangan Buah Sawit Berbasis Deep Learning dengan Menggunakan Arsitektur Yolov5. *Journal of Information System Research (JOSH)*, 5(1), 302–309. <https://doi.org/10.47065/josh.v5i1.4408>
- Martanto, M., & Bahtiar, A. (2022). Penerapan Algoritma *K-Means Clustering* Pada Pengelompokan Pasangan Usia Subur Peserta Kb Di Kabupaten Cirebon. *Jurnal Teknologi Technoscientia*, 15(1), 9–14. <https://doi.org/10.34151/technoscientia.v15i1.3849>
- Muntahanah, S., Arinastuti, A., & Budiastuti, S. (2024). Wirausaha sebagai upaya pengentasan kemiskinan. *Wikuacity: Jurnal Pengabdian Kepada Masyarakat*, 3(1), 103-106.
- Nazeriandy, N., Syahra, Y., & Syaifudin, M. (2021). Penerapan Data Mining Untuk Memprediksi Penggunaan Daya Listrik Pada PT. PLN (Persero) Rayon Medan Selatan Dengan Menggunakan Metode Regresi Linier Berganda. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, 20(1), 20-27.
- Qadrini, L. (2020). Metode *K-Means* dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017. 13(2), 5–11. <https://doi.org/10.36456/jstat.vol13.no2.a2886>
- Safitri, P. N., Aristawidya, R., & Faradilla, S. B. (2021). Klasterisasi Faktor-Faktor Kemiskinan Di Provinsi Jawa Barat Menggunakan K-Medoids Clustering. *Journal of Mathematics Education and Science*, 4(2), 75–80. <https://doi.org/10.32665/james.v4i2.242>
- Sari, D. N. P., & Sukestiyarno, Y. L. (2021, February). Analisis cluster dengan metode K-Means pada persebaran kasus COVID-19 berdasarkan Provinsi di Indonesia. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 4, pp. 602-610).
- Sari, F. D. R., & Ediwijoyo, S. P. (2021). Pemetaan Tingkat Kemiskinan Di Provinsi Jawa Tengah Berdasarkan Kabupaten/Kota Dengan Metode K-Medoids, *J. Media Inform. Budidarma*, 5(4), 1528.
- Sinatrya, N. S., & Wardhani, L. K. (2018, August). Analysis of K-Means and K-Medoids's Performance Using Big Data Technology. In *2018 6th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-5). IEEE.
- Syukron, H., Fauzi Fayyad, M., Junita Fauzan, F., Ikhsani, Y., & Rizky Gurning, U. (2022). Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 76 DAN

- Ulum, B., Tohidi, E., & Nuris, N. D. (2024). Penerapan Algoritma K-Medoids Dalam Menentukan Cluster Kabupaten/Kota Berdasarkan Migrasi Penduduk Jawa Barat. *Jurnal Teknik Informatika dan Teknologi Informasi*, 4(1), 111-122.
- Widianingsih, P., Darmawan, G., & Sunengsih, N. (2022). Analisis Intervensi dalam Model SARIMA untuk Memprediksi Laju Inflasi di Kota Tasikmalaya. *Formosa Journal of Science and Technology*, 1(4), 6. <https://doi.org/10.55927/fjst.v1i4.1030>
- World Bank. (2001). *World development report 2000/2001: Attacking poverty*. Oxford University Press. <https://doi.org/10.1596/0-1952-1119-5>
- Zahra, F., Khalif, A., Sari, B. N., Karawang, U. S., Timur, T., & Barat, J. (2024). Provinsi Indonesia Menggunakan Algoritma K-Medoids. 12(2).
- Zhu, A., Hua, Z., Shi, Y., Tang, Y., & Miao, L. (2021). An improved k-means algorithm based on evidence distance. *Entropy*, 23(11). <https://doi.org/10.3390/e23111550>