

MINI-BATCH K-MEANS CLUSTERING UNTUK PENGELOMPOKAN KEMANDIRIAN DAERAH DI SULAWESI SELATAN

Muh. Aidil Amri¹, Muh. Fajrin Bakri², Andi Akram Nur Risal³, Dewi Fatmarani Suriyanto⁴

^{1,2,3,4}Jurusan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Makassar, Indonesia
¹aidilamri2004@gmail.com, ²muhammadfajrinb@gmail.com, ³akramandi@unm.ac.id,
⁴dewifatmaranis@unm.ac.id

Abstrak

Kurangnya pemetaan yang efisien mengenai tingkat kemandirian desa di Sulawesi Selatan menghambat perumusan kebijakan pembangunan yang tepat sasaran. Penelitian ini bertujuan untuk mengembangkan model *clustering* menggunakan algoritma Mini-Batch K-Means guna mengelompokkan desa berdasarkan tingkat kemandiriannya. Data yang digunakan mencakup Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL) dari 2.000 desa di Sulawesi Selatan. Proses analisis melibatkan *pre-processing* data, penentuan jumlah kluster optimal menggunakan metode Elbow, serta penerapan *Mini-Batch K-Means* untuk pengelompokan desa. Hasil penelitian menunjukkan bahwa desa-desa dapat dikelompokkan ke dalam empat kluster utama: Desa Mandiri, Desa Maju, Desa Berkembang, serta gabungan Desa Tertinggal dan Desa Sangat Tertinggal. Evaluasi menggunakan Silhouette Coefficient mengindikasikan bahwa kluster Desa Mandiri memiliki kualitas pengelompokan terbaik dengan nilai 0,5398, sementara kluster Desa Berkembang memiliki nilai 0,1823, yang menunjukkan adanya tumpang tindih dengan kluster lain. Implementasi model ini memberikan gambaran yang lebih jelas terkait distribusi tingkat kemandirian desa dan dapat menjadi acuan dalam penyusunan strategi pembangunan yang lebih efektif dan berkelanjutan. Dari hasil penelitian ini, pemerintah daerah dan pemangku kepentingan dapat merumuskan kebijakan berbasis data yang lebih akurat, memastikan bahwa desa-desa yang membutuhkan intervensi lebih lanjut dapat memperoleh perhatian yang tepat. Rekomendasi strategi berbasis data ini diharapkan dapat meningkatkan efektivitas program pembangunan desa serta mempercepat pencapaian kesejahteraan yang merata di Sulawesi Selatan

Kata kunci : algoritma minibatch k-means, indeks desa membangun, kemandirian desa, *clustering*.

1. Pendahuluan

Sulawesi Selatan adalah provinsi di Indonesia yang memiliki keunikan dalam aspek sosial, budaya, dan ekonomi (Putri et al., 2024). Provinsi ini tidak hanya dikenal dengan keanekaragaman budaya dan kekayaan alamnya, tetapi juga dengan dinamika sosial ekonomi yang melibatkan desa-desa yang tersebar luas (Arifah et al., 2021). Desa-desa ini memainkan peran penting dalam pengembangan ekonomi dan pelestarian budaya, dengan berbagai inisiatif lokal yang melibatkan pengelolaan sumber daya alam serta tradisi budaya yang masih dilestarikan. Potensi yang besar ini seharusnya dimanfaatkan oleh desa-desa di Sulawesi Selatan untuk memperkuat kemandirian jangka panjang. Sayangnya, hingga kini belum ada pemetaan yang memadai untuk mengetahui desa-desa mana yang sudah mandiri dan yang belum (Arifin, 2024).

Hal ini dapat dibuktikan dari data Badan Pusat Statistik dimana hanya 17 desa yang tercatat sebagai desa mandiri pada tahun 2021 mencerminkan masih banyak desa yang belum mencapai status kemandirian atau bahkan belum terdata hingga saat ini (Badan Pusat Statistik, 2021). Kurangnya pemetaan ini dapat membuat beberapa desa merasa kurang mendapatkan perhatian dalam upaya

kemandirian. Tanpa data yang jelas, pemerintah dan pemangku kepentingan akan sulit mengidentifikasi desa yang memerlukan dukungan khusus untuk mencapai kemandirian. Situasi ini berpotensi menghambat perencanaan pembangunan yang berfokus dan berkelanjutan, serta mengurangi efektivitas program-program peningkatan ekonomi dan sosial di tingkat desa (Fita et al., 2022).

Sejumlah penelitian terkait pengelompokan desa berdasarkan kemandirian telah dilakukan dengan berbagai pendekatan. Sebagai contoh, pada penelitian menggunakan metode *Clustering* dan deskriptif untuk mengelompokkan desa-desa berdasarkan keberlanjutan di provinsi tertentu. Pendekatan ini menghasilkan kluster desa yang dapat dijadikan acuan untuk pengembangan ekonomi lokal melalui pendekatan kualitatif dan kuantitatif. Namun, cakupan variabel masih terbatas, sehingga perlu pengembangan lebih lanjut (Andria et al., 2023). Penelitian lain menerapkan Support Vector Machine (SVM) Multiclass pada data Indeks Desa Membangun (IDM) di Provinsi Maluku berhasil memperoleh akurasi sebesar 97,75%, namun membutuhkan biaya komputasi yang tinggi pada dataset yang besar (Latuconsina et al., 2024).

Selanjutnya, penelitian lainnya mengimplementasikan Fuzzy C-Means (FCM) untuk

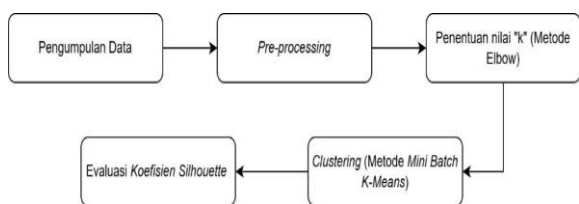
mengurangi subjektivitas dalam pengelompokan desa di Kalimantan Timur. Pendekatan ini menawarkan fleksibilitas dalam pengelompokan data desa, meskipun tidak optimal tanpa proses normalisasi yang baik (Syoyer & Wahyudin, 2021). Selain itu, penelitian lain melakukan klasifikasi desa di Kabupaten Landak dengan Algoritma C5.0. Hasilnya menunjukkan akurasi yang memadai dalam klasifikasi status desa, meski hanya mencakup sebagian variabel yang relevan (Nur Fitria et al., 2022). Penelitian lain menggunakan K-Means Clustering untuk clustering desa berdasarkan Indeks Desa Membangun (IDM). Metode ini terbukti efisien dalam analisis data skala besar namun kurang optimal dalam kasus distribusi kluster yang tidak merata (Harahap & Zulvia, 2021).

Dari beberapa penelitian sebelumnya berfokus dengan menggunakan metode K-Means atau SVM, sedangkan pada penelitian ini menggunakan algoritma pengembangan dari metode K-Means yaitu MiniBatch K-Means dalam melakukan pengelompokan desa berdasarkan kemandirian. Algoritma ini memiliki potensi untuk mengatasi kendala dalam analisis skala besar melalui pemrosesan yang lebih cepat. Salah satu tantangan utama dalam pemetaan kemandirian desa di Sulawesi Selatan adalah tidak adanya metode yang efisien untuk mengelompokkan desa secara tepat. Akibatnya, pemerintah dan pemangku kepentingan kesulitan mendapatkan data yang jelas untuk merumuskan kebijakan pembangunan desa.

Oleh karena itu, penelitian ini bertujuan mengembangkan model clustering dengan metode MiniBatch K-Means yang mampu memetakan jelas mengenai desa-desa yang sudah mandiri maupun yang memerlukan intervensi khusus di Sulawesi Selatan. Dengan adanya penelitian ini, diharapkan dapat membantu pemerintah daerah dan pemangku kepentingan dalam mengidentifikasi dan memetakan desa-desa di Sulawesi Selatan berdasarkan tingkat kemandiriannya. Dengan demikian, perencanaan pembangunan desa dapat dilakukan secara lebih tepat sasaran dan berkelanjutan, memperkuat kemandirian ekonomi dan sosial desa dalam jangka panjang.

2. Metode

Metodologi dalam penelitian ini terdiri dari beberapa tahapan yang diselesaikan secara berurutan. Tahapan-tahapan tersebut meliputi pengumpulan data, *pre-processing*, implementasi algoritma, klasifikasi dan evaluasi yang disajikan pada Gambar 1.



Gambar 1. Tahapan metode penelitian

2.1 Pengumpulan Data

Pada tahap ini, dilakukan pengumpulan data berupa indeks status kemajuan dan kemandirian desa seluruh Indonesia. Data ini diperoleh dari Satu Data Indonesia pada tahun 2024 dengan total data sebanyak 75,425 data. Data ini memiliki beberapa atribut yaitu kode prov, nama provinsi, kode kab, nama kabupaten, kode kec, nama kecamatan, nama desa, Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), Indeks Ketahanan Lingkungan (IKL), dan nilai Indeks Desa Membangun (IDM).

Proses pengumpulan data dilakukan dengan mengakses portal resmi *Satu Data Indonesia* yang menyediakan informasi terkait indeks pembangunan desa. Data diunduh dalam format CSV dan kemudian dikonversi ke dalam format yang sesuai untuk analisis lebih lanjut. Setelah itu, dilakukan proses verifikasi dan validasi data dengan membandingkan hasil unduhan dengan data tahun sebelumnya guna memastikan kelengkapan dan konsistensi informasi yang digunakan.

2.2 Pre-processing

Setelah dilakukan pengumpulan data, tahap selanjutnya adalah *pre-processing*. Data mentah seringkali mengandung masalah seperti data yang tidak lengkap, inkonsistensi format, nilai duplikat, atau bahkan data yang tidak relevan dengan tujuan penelitian (Tri & Nataliani, 2021). *Pre-processing* dilakukan untuk meningkatkan kualitas data, memastikan konsistensi, dan mengurangi noise atau informasi yang tidak diperlukan. Pada penelitian ini dilakukan sejumlah tahapan *pre-processing*. Langkah pertama dalam *preprocessing* adalah pembersihan data (*data cleaning*), di mana data diperiksa untuk mendeteksi dan menghapus nilai duplikat, nilai kosong, atau data yang tidak relevan guna memastikan kualitas data yang lebih baik. Selain itu, dilakukan pengurangan dimensi untuk mengatasi kompleksitas dataset dengan mempertahankan atribut yang memiliki signifikansi tinggi dalam analisis clustering. Pemilihan kolom yang dikurangi dilakukan berdasarkan analisis korelasi, dengan tujuan menghilangkan atribut yang tidak memberikan kontribusi signifikan terhadap hasil clustering. Kolom yang dikurangi meliputi kode provinsi, kode kabupaten, kode kecamatan, dan kode desa karena informasi tersebut bersifat administratif dan tidak berpengaruh langsung terhadap variabel utama. Variabel utama yang digunakan dalam penelitian ini, yaitu IKS, IKE, IKL, dan IDM, dipilih karena memiliki relevansi yang tinggi dalam mengukur tingkat kemandirian desa.

Kolom-kolom yang dihapus tidak digunakan dalam pengukuran tingkat kemandirian desa karena hanya berfungsi sebagai identitas wilayah dan tidak memberikan nilai informatif dalam analisis clustering. Setelah tahap pembersihan data selesai, dilakukan proses *filtering* untuk memastikan bahwa

hanya data desa di Provinsi Sulawesi Selatan yang digunakan dalam analisis. Proses ini dilakukan dengan menyaring data berdasarkan atribut nama provinsi, sehingga desa-desa dari provinsi lain dihilangkan dari dataset. Penyaringan ini penting karena penelitian ini berfokus pada *clustering* tingkat kemandirian desa di Sulawesi Selatan, sehingga data dari provinsi lain tidak relevan untuk analisis yang dilakukan.

2.3 Dataset

Dapat dilihat pada Tabel 1 bahwa data tersebut belum dilakukan preprocessing, di mana dataset masih berupa indeks dari seluruh desa di Indonesia. Selanjutnya, pada Tabel 2 dapat dilihat data yang telah dilakukan preprocessing, dataset ini difilter untuk difokuskan pada desa-desa yang terletak di Provinsi Sulawesi Selatan saja. Proses filtering dilakukan berdasarkan atribut "Nama Provinsi," sehingga menghasilkan subset data yang relevan untuk penelitian ini. Selain itu, beberapa kolom yang tidak diperlukan dihapus setelah dilakukan tahap *preprocessing*.

2.4 Metode Elbow

Metode Elbow adalah pendekatan dalam analisis kluster yang digunakan untuk menentukan jumlah optimal kluster dalam suatu dataset. Metode ini bekerja dengan menghitung fungsi biaya yang disebut *cost function* (Liu & Deng, 2021). Metode Elbow merupakan teknik evaluasi dalam algoritma *Clustering*, khususnya untuk menentukan jumlah kluster optimal (*k*) dalam analisis K-Means. Metode ini bekerja dengan memplot nilai *Within-Cluster Sum of Squares* (WCSS) terhadap berbagai nilai *k*. WCSS menghitung jumlah total dari kuadrat jarak antara setiap titik data dan centroid klasternya masing-masing (Thongnim et al., 2024). Titik "elbow" adalah saat penurunan nilai WCSS mulai melambat, yang menunjukkan bahwa menambahkan lebih banyak kluster tidak signifikan meningkatkan representasi data. Titik ini dipilih sebagai *k* yang optimal. Adapun rumus dari metode *elbow* dapat dilihat pada persamaan 1 (Lashiyanti et al., 2023):

$$WCSS = \sum_{k=1}^K \sum_{i \in C_k} |x_i - \mu_k|^2 \tag{1}$$

Keterangan:

k = Jumlah kluster.

C_k = Kluster ke-*k*.

x_i = Titik data ke- *i* dalam kluster *C_k*.

μ_k = Centroid kluster *k*.

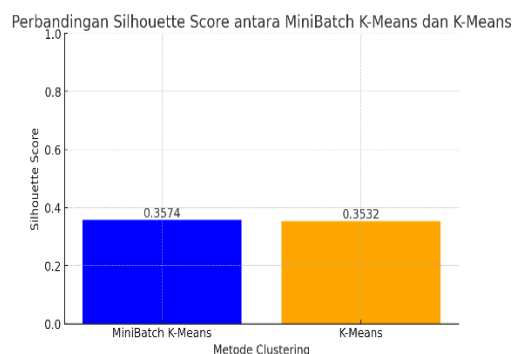
$|x_i - \mu_k|^2$ = Jarak Euclidean kuadrat antara titik data *x_i* dan centroid *μ_k*.

Nilai WCSS dihitung untuk berbagai jumlah kluster (*k*) yang diuji, yaitu dari *k*=2 hingga *k*=10. Grafik WCSS kemudian diplot untuk menemukan titik "elbow" yaitu jumlah kluster di mana penurunan nilai WCSS mulai melambat. Berdasarkan

hasil analisis, jumlah kluster optimal ditemukan pada *k*=4, yang sesuai dengan kategori desa yang telah ditentukan.

2.5 Mini-batch K-Means

Mini-batch k-means adalah algoritma *Clustering* yang dirancang untuk menangani dataset besar dengan memori terbatas (Hicks et al., 2021). Setelah jumlah kluster optimal *k*=4 ditentukan, algoritma *MiniBatch K-Means* digunakan untuk melakukan *Clustering*. MiniBatch K-Means dipilih dalam penelitian ini karena memiliki efisiensi yang lebih tinggi dibandingkan K-Means biasa dalam menangani dataset skala besar. Berdasarkan pada Gambar 2, dilakukan perbandingan nilai Silhouette Score untuk MiniBatch K-Means (0.3574) sedikit lebih tinggi dibandingkan dengan K-Means biasa (0.3532), yang menunjukkan bahwa metode ini tetap memberikan hasil *clustering* yang baik dengan waktu pemrosesan yang lebih cepat. Penggunaan *mini-batches* dalam MiniBatch K-Means juga memungkinkan algoritma untuk mengolah data dengan konsumsi memori yang lebih rendah, menjadikannya pilihan yang lebih optimal untuk penelitian ini.



Gambar 2. Nilai Silhouette Score MiniBatch K-Means dan K-Means

Adapun langkah-langkah dari dari algoritma sebagai berikut (Mulyadi et al., 2024) :

- a. Subset data (mini-batch) diambil secara acak dari dataset.
- b. Jarak Euclidean (Mulyadi et al., 2024): Perhitungan dengan jarak Euclidean dapat dilihat pada persamaan 2

$$d(x_i, \mu_k) = \sqrt{\sum_{j=1}^m (x_{ij} - \mu_{kj})^2} \tag{2}$$

Keterangan:

x_i = Vektor data ke-*i*.

μ_k = Centroid kluster ke- *k*.

m = Jumlah dimensi (fitur) data.

- c. Pemetaan Data ke Kluster (Mulyadi et al., 2024):

Rumus untuk perhitungan untuk pemetaan data ke kluster pada persamaan 3

$$C_k = \arg_k \min d(x_i, \mu_k) \tag{3}$$

Keterangan :

Data x_i = ditugaskan ke kluster dengan jarak Euclidean terkecil.

- d. Pembaruan Centroid (Mini-Batch) (Mulyadi et al., 2024) :

Rumus untuk pembaruan centroid dengan *mini-batch* pada persamaan 4

$$\mu_k^{new} = \mu_k^{old} + \eta \cdot (x_i - \mu_k^{old}) \tag{4}$$

Keterangan :

μ_k^{new} = Centroid baru kluster k .

μ_k^{old} = Centroid sebelumnya.

x_i = Titik data dari mini-batch yang ditugaskan ke kluster k .

η = *Learning rate* atau faktor pembaruan

2.6 Silhouette Coefficient

Silhouette coefficient adalah metrik evaluasi untuk mengukur seberapa baik suatu *Cluster* dalam data. Matrik ini digunakan untuk mengevaluasi kualitas kluster dalam analisis *clustering* (Punhani et al., 2022). *Silhouette Coefficient* memiliki nilai antara -1 dan 1, di mana nilai positif menunjukkan bahwa objek dalam sebuah *Cluster* lebih dekat satu sama lain dibandingkan ke *Cluster* lain, sementara nilai negatif menunjukkan bahwa objek tersebut mungkin lebih sesuai dengan *Cluster* lain. Berikut adalah langkah perhitungan *silhouette coefisien* (Mulyadi et al., 2024):

- 3.1 Hitung rata-rata jarak antara objek dan semua objek lainnya dalam *Cluster*. Untuk menghitung jarak antara objek dan semua objek pada persamaan 5.

$$va(i) = 1 | A | - 1 \sum_{j \in A, j \neq i} d(i, j) a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \tag{5}$$

Keterangan:

A = jumlah data dalam *Cluster*

- 3.2 Hitung rata-rata jarak objek ke semua objek lainnya dalam *Cluster* lain, kemudian gunakan persamaan 6 untuk mendapatkan nilai terkecil.

$$nd(i, C) = 1 | C | \sum_{j \in A} d(i, j) d(i, C) = \frac{1}{|C|} \sum_{j \in A} d(i, j) \tag{6}$$

Keterangan:

C = jumlah data dalam *Cluster* C

- 3.3 Hitung nilai koefisien Silhouette dengan persamaan 7.

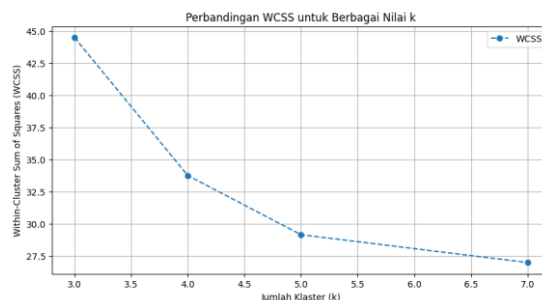
$$ms(i) = b(i) - a(i) \max a(i), b(i) s(i) = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}} \tag{7}$$

Nilai $s(i)$ berkisar antara -1 dan 1, di mana setiap nilai diinterpretasikan sebagai berikut:

- 1) $s(i) \approx 1$ data i terkelompok dengan baik (dalam A)
- 2) $s(i) \approx 0$ data i berada ditengah tengah antara dua *Cluster* (A dan B)
- 3) $s(i) \approx -1$ data i diklasifikasikan sebagai lemah (lebih dekat ke *Cluster* B daripada A)

3. Hasil dan Pembahasan

Pada penelitian ini, data yang digunakan memiliki rentang nilai antara 0 hingga 1, sehingga proses normalisasi data tidak diperlukan. Hal ini dikarenakan rentang nilai yang seragam memungkinkan data untuk langsung digunakan dalam proses *Clustering* tanpa memengaruhi hasil pengelompokan. Untuk menentukan jumlah kluster yang optimal, digunakan metode Elbow, yang divisualisasikan melalui grafik pada Gambar 2. Grafik Elbow ini menunjukkan hubungan antara jumlah kluster (k) dengan nilai *Within-Cluster Sum of Squares* (WCSS), yang merepresentasikan total jarak antara data dalam setiap kluster terhadap centroidnya.



Gambar 3. Nilai WCSS Terhadap Jumlah K

Berdasarkan hasil analisis menggunakan Metode Elbow dan Silhouette Score, dapat dilihat pada Gambar 3 jumlah cluster yang optimal dalam penelitian ini adalah $k=4$. Metode Elbow menunjukkan adanya titik perubahan signifikan pada $k=4$, di mana setelah titik ini penurunan nilai *Within-Cluster Sum of Squares* (WCSS) mulai melandai. Hal ini menandakan bahwa penambahan jumlah cluster setelah $k=4$ tidak lagi memberikan pengurangan variansi yang signifikan dalam masing-masing cluster, sehingga jumlah cluster ini dianggap sebagai titik optimal dalam keseimbangan antara variasi intra-cluster dan kompleksitas model. Selain itu pada Tabel 2 disajikan hasil perhitungan Silhouette Score yang menunjukkan bahwa nilai tertinggi diperoleh pada $k=4$ dengan skor 0.357 dibandingkan jumlah cluster lainnya ($k=3, k=5, k=7$), yang mengindikasikan bahwa cluster yang terbentuk pada $k=4$ memiliki pemisahan yang lebih jelas serta tingkat kepadatan yang lebih baik.

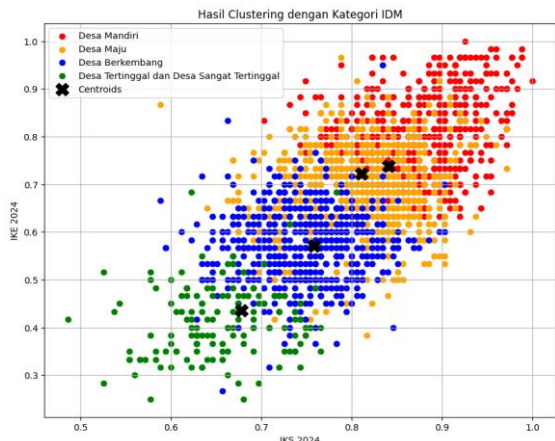
Tabel 1. Perbandingan Kinerja Clustering

Jumlah Kluster (K) Mini Batch K-Means	Silhouette Score	Nilai WCSS
3	0.3422	44.65
4	0.3574	33.48
5	0.3104	29.53
7	0.2806	23.02

Dengan mempertimbangkan hasil dari *Silhouette Score* dan Nilai WCSS, k=4 dipilih sebagai jumlah cluster yang optimal karena memberikan keseimbangan terbaik antara homogenitas dalam cluster (cohesion) dan perbedaan antar cluster (separation), sehingga dapat menghasilkan pengelompokan yang lebih representatif dan bermakna.

Oleh karena itu, jumlah kluster optimal yang dipilih adalah empat, yang sesuai dengan kategori IDM, yaitu Desa Mandiri, Desa Maju, Desa Berkembang, serta gabungan Desa Tertinggal dan Desa Sangat Tertinggal.

Setelah jumlah kluster optimal ditentukan menggunakan metode Elbow, proses pengelompokan desa-desa dilakukan menggunakan algoritma MiniBatch K-Means. Hasil dari proses ini divisualisasikan pada Gambar 2, yang menunjukkan distribusi desa berdasarkan nilai IKS dan IKE. Nilai IKS dan IKE dipilih untuk divisualisasikan karena keduanya secara langsung merepresentasikan dua dimensi penting dalam pengelompokan desa.



Gambar 4. Sebaran Hasil Clustering dengan Mini-Batch K-Means

Grafik pada Gambar 4. menunjukkan hasil pengelompokan desa-desa di Sulawesi Selatan berdasarkan kategori IDM menggunakan metode MiniBatch K-Means. Pengelompokan dilakukan ke dalam empat kategori utama: Desa Mandiri (merah), Desa Maju (kuning), Desa Berkembang (biru), serta gabungan Desa Tertinggal dan Desa Sangat Tertinggal (hijau). Distribusi data menunjukkan bahwa desa dengan nilai IDM lebih tinggi untuk Desa Mandiri cenderung berada di bagian kanan atas grafik, sementara desa dengan nilai IDM lebih rendah untuk Desa Tertinggal dan Desa Sangat Tertinggal berada di bagian kiri bawah grafik.

Namun, terdapat kedekatan centroid antara kluster desa mandiri dan desa maju. Hal ini dikarenakan rentang nilai IDM untuk desa mandiri dan desa maju relatif berdekatan. Selain itu, indikator sosial, ekonomi dan lingkungan pada kedua kategori ini memiliki kesamaan yang cukup signifikan yang menyebabkan beberapa desa dalam kategori ini memiliki karakteristik yang hampir serupa, sehingga memengaruhi posisi centroid yang menjadi lebih dekat. Meski demikian, hasil *Clustering* tetap valid karena mempertimbangkan nilai IDM sebagai dasar kategorisasi yang utama.

3.1 Kluster 1

Pada kluster 1 mencakup Desa Mandiri sebanyak 744 desa. Untuk merepresentasikan kluster ini, diambil 4 desa dengan nilai *silhouette score* tertinggi sebagai sampel. Data pada Tabel 3 menunjukkan bahwa Desa Corawali mencatatkan nilai *silhouette score* sebesar 0,6316, dengan nilai IKS sebesar 0,8457, IKE sebesar 0,7500, dan IKL sebesar 0,9333, menghasilkan skor IDM 0,8430. Desa ini menunjukkan tingkat kemandirian yang sangat tinggi dalam aspek sosial, ekonomi, dan lingkungan.

Di sisi lain, desa dengan nilai *silhouette score* terendah dalam kluster ini adalah Desa Bukit Tinggi dengan nilai *silhouette score* sebesar 0,0361, meskipun memiliki nilai IDM sebesar 0,8321. Hal ini menunjukkan bahwa meskipun desa ini termasuk dalam kategori Desa Mandiri, karakteristiknya cenderung mendekati batas kategori lain, yang mengindikasikan adanya tumpang tindih dalam aspek sosial, ekonomi, atau lingkungan.

3.2 Kluster 2

Pada *Cluster* Desa Maju, terdapat total 815 desa, dengan 5 desa diambil sebagai contoh berdasarkan nilai *Silhouette Score* tertinggi. Dapat dilihat pada Tabel 5 salah satu desa yang memiliki nilai *silhouette score tertinggi* adalah Desa Wellulang dengan nilai sebesar 0.5558. Desa ini memiliki nilai IKS sebesar 0,7771, IKE sebesar 0,7333, dan IKL sebesar 0,9333, menghasilkan nilai IDM sebesar 0,8146. Secara sosial, ekonomi, dan lingkungan, desa ini menunjukkan perkembangan signifikan menuju kemandirian.

Namun, terdapat desa dengan nilai *Silhouette Score* yang rendah, seperti Desa Manjalling dengan nilai *silhouette score* sebesar -0,0723, meskipun memiliki nilai IDM sebesar 0,7078. Ini mengindikasikan bahwa desa tersebut berada lebih dekat dengan centroid kluster lain dibandingkan dengan klasternya sendiri, yaitu Desa Maju. Ini menunjukkan bahwa pengelompokan desa ini mungkin kurang sesuai, dan karakteristiknya lebih mirip dengan desa-desa pada kluster yang berbeda.

3.3 Kluster 3

Kluster ini merupakan kategori desa berkembang mencakup sebanyak 567 desa dengan nilai IDM

antara 0,5989 hingga 0,7072 yang dapat dilihat pada Tabel 6. Diambil 5 desa sebagai contoh berdasarkan nilai *silhouette score* tertinggi. Salah satunya adalah Desa Kaduaja yang memiliki nilai IKS sebesar 0,8000, IKE sebesar 0,7167, dan IKL sebesar 0,6000, menghasilkan nilai IDM sebesar 0,7056 dengan nilai *silhouette score* tertinggi sebesar 0,4745. Hal ini menunjukkan bahwa desa ini sesuai dengan karakteristik *Cluster* Desa Berkembang, yang mencerminkan tahap transisi menuju kemandirian yang lebih baik.

Sedangkan, terdapat beberapa desa dengan nilai *silhouette score* yang rendah bahkan negatif, seperti Desa Lampulung dengan nilai *silhouette score* sebesar -0,1045. Nilai IDM desa ini sebesar 0,6311, yang berada di dekat batas bawah kategori Desa Berkembang yaitu 0,5989 hingga 0,7072. Hal ini menunjukkan bahwa karakteristik desa ini sedikit mendekati *Cluster* Desa Tertinggal, meskipun secara teknis masih termasuk dalam kategori Desa Berkembang.

3.4 Kluster 4

Kluster ini mencakup sebanyak 140 desa, dengan nilai IDM antara 0,4907 hingga 0,5989 yang dapat dilihat pada Tabel 7. Pada kluster ini, merupakan gabungan dari kategori Desa Tertinggal dan Desa Sangat Tertinggal, yang semula terpisah. Penggabungan ini dilakukan karena jumlah data pada kategori Desa Sangat Tertinggal sangat sedikit, sehingga lebih relevan untuk digabungkan dengan kategori Desa Tertinggal guna memberikan hasil analisis yang lebih komprehensif. Contohnya adalah Desa Basseang memiliki nilai IKS sebesar 0,6629, IKE sebesar 0,4000, dan IKL sebesar 0,6000, menghasilkan nilai IDM sebesar 0,5543, dengan *silhouette score* tertinggi sebesar 0,5380. Desa ini menunjukkan pengelompokan yang sangat sesuai dengan karakteristik *Cluster* ini, mencerminkan desa yang memerlukan pengembangan lebih lanjut untuk meningkatkan taraf kemandirian.

Namun, pada data terakhir terdapat Desa Bukit Sutra yang sebelumnya termasuk dalam kategori Desa Sangat Tertinggal dengan nilai IDM 0,5910. Desa ini memiliki nilai *silhouette score* sebesar 0,0349, yang menunjukkan adanya kedekatan dengan *Cluster* lain. Meskipun secara teknis masuk ke dalam kategori Desa Tertinggal, desa ini tetap memerlukan evaluasi lebih lanjut karena karakteristiknya yang berada di batas atas kategori ini.

Hasil pengelompokan desa-desa di Sulawesi Selatan menggunakan algoritma MiniBatch K-Means memberikan gambaran yang jelas tentang tingkat kemandirian desa berdasarkan Indeks Desa Membangun (IDM). Desa-desa dikelompokkan ke dalam empat kategori utama, yaitu Desa Mandiri, Desa Maju, Desa Berkembang, serta Desa Tertinggal dan Desa Sangat Tertinggal. Evaluasi dengan *silhouette coefficient* menunjukkan bahwa metode

pengelompokan yang digunakan dapat memisahkan data dengan baik ke dalam kluster masing-masing.

Setelah pengelompokan dilakukan dan kategori desa ditentukan, analisis lebih lanjut dilakukan untuk mengevaluasi kualitas hasil *Clustering*. Hal ini dilakukan dengan menghitung rata-rata nilai *silhouette coefficient* untuk setiap kluster, yang memberikan gambaran seberapa baik data dalam kluster terpisah dari kluster lainnya. Hasil rata-rata nilai *silhouette coefficient* disajikan pada Tabel 8.

Tabel 2. Silhouette Coefficient Hasil Kluster

Kluster	Nilai Silhouette	Keterangan
Kluster 1	0,539	Baik
Kluster 2	0,315	Cukup Baik
Kluster 3	0,182	Kurang Baik
Kluster 4	0,428	Cukup Baik

Perbedaan yang signifikan terlihat di antara keempat kluster yang terbentuk ketika kualitas pengelompokan dinilai dengan menggunakan nilai *silhouette coefficient*. *Cluster* 1 memiliki nilai rata-rata *silhouette* tertinggi sebesar 0,5398, yang mengindikasikan bahwa desa-desa dalam kluster ini terkelompok dengan baik dan memiliki jarak yang jelas dari kluster lain. *Cluster* 2 memiliki nilai rata-rata Silhouette sebesar 0,3152, yang juga tergolong cukup baik, tetapi kualitas pengelompokannya lebih rendah dibanding *Cluster* 4. Sementara itu, *Cluster* 3 memiliki nilai rata-rata *silhouette* terendah sebesar 0,1823, yang mengindikasikan bahwa pengelompokan untuk kluster ini kurang baik. Desa-desa dalam *Cluster* 3 kemungkinan memiliki tumpang tindih dengan kluster lain. Kemudian dengan nilai Silhouette sebesar 0,4287, Kluster 4 berada pada tingkat cukup baik, yang berarti bahwa meskipun beberapa desa mungkin dekat dengan batas-batas kluster, sebagian besar desa di kluster ini cukup terpisah dari kluster lainnya.

4. Kesimpulan

Penelitian ini bertujuan untuk mengembangkan model *clustering* yang dapat memetakan desa-desa di Sulawesi Selatan berdasarkan tingkat kemandirian mereka. Dengan menggunakan algoritma MiniBatch K-Means, penelitian ini berhasil mengelompokkan 2,000 desa ke dalam empat kategori yaitu Desa Mandiri, Desa Maju, Desa Berkembang, serta gabungan Desa Tertinggal dan Desa Sangat Tertinggal. Proses ini dimulai dengan pengumpulan data dari Satu Data Indonesia, yang mencakup berbagai indeks kemajuan dan kemandirian desa.

Metode Elbow digunakan untuk menentukan jumlah kluster optimal, yang ditemukan pada k=4. Hasil pengelompokan menunjukkan bahwa desa-desa dengan nilai Indeks Desa Membangun (IDM) yang lebih tinggi cenderung berada dalam kategori Desa Mandiri, sedangkan desa dengan nilai IDM yang lebih rendah termasuk dalam kategori Desa Tertinggal dan Desa Sangat Tertinggal. Evaluasi menggunakan Silhouette Coefficient menunjukkan

bahwa kualitas pengelompokan bervariasi antar kluster, dengan Kluster 1 (Desa Mandiri) memiliki nilai Silhouette tertinggi (0,5398), menunjukkan pengelompokan yang baik, sedangkan Kluster 3 (Desa Berkembang) memiliki nilai terendah (0,1823), menunjukkan adanya tumpang tindih dengan kluster lain.

Meskipun penelitian ini berhasil memetakan desa-desa berdasarkan tingkat kemandiriannya, terdapat keterbatasan yaitu metode *MiniBatch K-Means* yang digunakan memiliki keterbatasan dalam menangani distribusi kluster yang tidak merata, yang menyebabkan beberapa kluster memiliki nilai silhouette rendah dan terjadi tumpang tindih antara kluster. Untuk penelitian selanjutnya, disarankan agar dilakukan eksplorasi terhadap metode clustering yang lebih kompleks, seperti DBSCAN atau Gaussian Mixture Model (GMM), yang dapat menangani distribusi data yang lebih bervariasi. Selain itu, implementasi metode ensemble clustering dapat dipertimbangkan untuk meningkatkan akurasi dalam pemetaan kemandirian desa.

Hasil penelitian ini diharapkan dapat membantu pemerintah daerah dan pemangku kepentingan dalam merumuskan kebijakan pembangunan yang lebih tepat sasaran dan berkelanjutan, serta meningkatkan kemandirian ekonomi dan sosial desa di Sulawesi Selatan. Penelitian ini juga membuka peluang untuk penelitian lanjutan dalam pengembangan model yang lebih kompleks dan akurat dalam pengelompokan desa.

Daftar Pustaka:

- Andria, F., Rahmi, A., & Gunawan, A. D. (2023). IDM Clustering: Study of Village Potential Development Efforts in the Developing Category. *International Journal of Latest Research in Humanities and Social Science (IJLRHSS)*, 06(01), 249–257.
- Arifah, N., Anggraeni, R., & Mangilep, A. U. A. (2021). Studi Implementasi Sistem Rujukan Berjenjang Antar Fasilitas Kesehatan Tingkat Lanjut di Era JKN di Provinsi Sulawesi Selatan. *Jurnal Manajemen Kesehatan Indonesia*, 9(1), 45–52. <https://doi.org/10.14710/jmki.9.1.2021.45-52>
- Arifin, F. (2024). Optimalisasi Pengelolaan Keuangan dan Aset Desa dalam Rangka Meningkatkan Efektivitas Pemerintahan Desa. *Jurnal Ilmiah Hukum dan Hak Asasi Manusia*, 4(1), 35–46. <https://doi.org/10.35912/jihham.v4i1.3397>
- Badan Pusat Statistik. (2021). *Jumlah Desa Mandiri Menurut Provinsi (Desa), 2019-2021*. Jumlah Desa Mandiri. <https://www.bps.go.id/id/statistics-table/2/MjE5MCMY/jumlah-desa-mandiri-menurut-provinsi.html>
- Fita, G. A., Asriani, A., Amiruddin, A., & Fariaty, C. N. (2022). Pendampingan Analisis Potensi Desa Wisata Di Desa Pamboborang Kecamatan Banggae Kabupaten Majene. *MINDA BAHARU*, 6(1), 134–143. <https://doi.org/10.33373/jmb.v6i1.3909>
- Harahap, A. S., & Zulvia, P. (2021). Clustering Desa dengan Menggunakan Algoritma K-Means pada Data Potensi Desa. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 237. <https://doi.org/10.30865/jurikom.v8i6.3724>
- Hicks, S. C., Liu, R., Ni, Y., Purdom, E., & Risso, D. (2021). mbkmeans: Fast clustering for single cell data using mini-batch k-means. *PLOS Computational Biology*, 17(1), e1008625. <https://doi.org/10.1371/journal.pcbi.1008625>
- Lashiyanti, A. R., Munthe, I. R., & Nasution, F. A. (2023). Optimisasi Clustering Nilai Ujian Nasional dengan Pendekatan Algoritma K-Means, Elbow, dan Silhouette. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 6(1), 14–20.
- Latuconsina, F., Noya Van Delsen, M. S., & Yudistira. (2024). Klasifikasi Menggunakan Metode Support Vector Machine (SVM) Multiclass pada Data Indeks Desa Membangun (IDM) di Provinsi Maluku. *Journal of Mathematics, Computations and Statistics*, 7(2), 380–395. <https://doi.org/10.35580/jmathcos.v7i2.3624>
- Liu, F., & Deng, Y. (2021). Determine the Number of Unknown Targets in Open World Based on Elbow Method. *IEEE Transactions on Fuzzy Systems*, 29(5), 986–995. <https://doi.org/10.1109/TFUZZ.2020.2966182>
- Mulyadi, S., Insani, F., Agustian, S., & Afriyanti, L. (2024). Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mini Batch K-Means Clustering: Grouping Electricity Distribution Data Using The Mini Batch K-Means Clustering Algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1051–1062. <https://doi.org/10.57152/malcom.v4i3.1425>
- Nur Fitria, M. C., Debatara, N. N., & Rizki, S. W. (2022). Classification of Village Status in Landak Regency Using C5.0 Algorithm. *Tensor: Pure and Applied Mathematics Journal*, 3(1), 33–42. <https://doi.org/10.30598/tensorvol3iss1pp33-42>
- Punhani, A., Faujdar, N., Mishra, K. K., & Subramanian, M. (2022). Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means. *IEEE Access*, 10, 115025–115032.

- <https://doi.org/10.1109/ACCESS.2022.3215568>
- Putri, C. J. A., Wahyudi, B., Utama, A. P., Widodo, P., & Saragih, H. J. R. (2024). Analisis Budaya Damai Suku To Balo dan Masyarakat Desa Bulu Bulu Kabupaten Barru Provinsi Sulawesi Selatan. *Jurnal Kewarganegaraan*, 8(1), 1229–1238. <https://doi.org/10.31316/jk.v8i1.6452>
- Syoer, R. R., & Wahyudin, Y. (2021). (CLUSTER ANALYSIS WITH FUZZY CLUSTERING ALGORITHM Case Study Grouping Villages in Kalimantan Timur Province). *BESTARI: Buletin Statistik dan Aplikasi Terkini*, 1(1).
- Thongnim, P., Sreekajon, J., & Pukseng, T. (2024). Enhancing Durian Cultivation Efficiency Through Data-Driven Smart Farming Using Cluster Analysis and Machine Learning. *2024 5th International Conference on Big Data Analytics and Practices (IBDAP)*, 67–72. <https://doi.org/10.1109/IBDAP62940.2024.10689710>
- Tri, M. F., & Nataliani, Y. (2021). Analisis Pengaruh Penilaian Asesor terhadap Kinerja Guru Mata Pelajaran dengan k-Means Clustering. *Indonesian Journal of Computing and Modeling*, 4(1), 14–22. <https://doi.org/10.24246/icm.v4i1.5063>

Daftar Tabel

Tabel 3. Data sebelum pre-processing

Kode Prov	Nama Provinsi	Kode Kab	Nama Kabupaten	Kode Kec	Nama Kecamatan	Nama Desa	IKS 2024	IKE 2024	IKL 2024	IDM 2024
11	Aceh	1101	Aceh Selatan	110101	Bakongan	Keude Bakongan	0,8000	0,9000	0,5333	0,7444
11	Aceh	1101	Aceh Selatan	110101	Bakongan	Ujong Mangki	0,6629	0,6333	0,5333	0,6098
11	Aceh	1101	Aceh Selatan	110101	Bakongan	Gampong Drien	0,6743	0,5333	0,6000	0,6025
...
96	Papua Barat Daya	9610	Maybrat	961024	Mare Selatan	Sire Timur	0,5600	0,2500	0,9333	0,5811

Tabel 4. Data Setelah Pre-processing

Nama Provinsi	Nama Kabupaten	Nama Kecamatan	Nama Desa	IKS 2024	IKE 2024	IKL 2024	IDM 2024
Sulawesi Selatan	Kepulauan Selayar	Bontoharu	Bontoborusu	0,6971	0,6333	0,7333	0,6879
Sulawesi Selatan	Kepulauan Selayar	Bontoharu	Bontolebang	0,6229	0,5833	0,6667	0,6243
Sulawesi Selatan	Bantaeng	Sinoa	Pallantikang	0,7943	0,6833	0,6667	0,7148
...	0,8597
Sulawesi Selatan	Gowa	Pallangga	Taeng	0,8743	0,6333	1,0000	0,8359

Tabel 5. Hasil Kluster Desa Mandiri

Nama Desa	IKS 2024	IKE 2024	IKL 2024	IDM 2024	Kategori Idm	Silhouette Score
Corawali	0,8457	0,75	0,933	0,843	Desa Mandiri	0,631
Lowa	0,8457	0,75	0,933	0,843	Desa Mandiri	0,631
Siawung	0,84	0,75	0,933	0,841	Desa Mandiri	0,631
Mallasoro	0,8343	0,75	0,933	0,839	Desa Mandiri	0,630
...
Bukit Tinggi	0,8629	0,9	0,7333	0,8321	Desa Mandiri	0,0361

Tabel 7. Hasil Kluster Desa Maju

Nama Desa	IKS 2024	IKE 2024	IKL 2024	IDM 2024	Kategori IDM	Silhouette Score
Wellulang	0,777	0,733	0,933	0,814	Desa Maju	0,555
Tocinnong	0,805	0,7	0,933	0,813	Desa Maju	0,550
Bonto Padang	0,782	0,716	0,933	0,811	Desa Maju	0,543
Laponrong	0,828	0,683	0,933	0,815	Desa Maju	0,543
...
Manjalling	0,84	0,616	0,666	0,707	Desa Maju	-0,072

Tabel 6. Hasil Kluster Desa Berkembang

Nama Desa	IKS 2024	IKE 2024	IKL 2024	IDM 2024	Kategori IDM	Silhouette Score
Kaduaja	0,8	0,716	0,6	0,705	Desa Berkembang	0,474
Buntu Tangti	0,765	0,75	0,6	0,705	Desa Berkembang	0,466
Buntu Kunyi	0,771	0,733	0,6	0,701	Desa Berkembang	0,457
Belau Utara	0,782	0,533	0,8	0,705	Desa Berkembang	0,451
...
Lampulung	0,76	0,5333	0,6	0,6311	Desa Berkembang	-0,1045

Tabel 8. Hasil Kluster Desa Tertinggal dan Sangat Tertinggal

Nama Desa	Iks 2024	Ike 2024	Ikl 2024	Idm 2024	Kategori Idm	Silhouette Score
Basseang	0,662	0,4	0,6	0,554	Desa Tertinggal	0,537
Pangi	0,662	0,4	0,6	0,554	Desa Tertinggal	0,537
Makkodo	0,64	0,416	0,6	0,552	Desa Tertinggal	0,536
Siteba	0,674	0,4167	0,6	0,563	Desa Tertinggal	0,535
...
Bukit Sutra	0,622	0,683	0,466	0,591	Desa Sangat Tertinggal	0,034