Arus Jurnal Sosial dan Humaniora (AJSH)
Website: http://jurnal.ardenjaya.com/index.php/ajsh
Email: jurnal.ajsh@ardenjaya.com

Pengelompokan Desa/Kelurahan di Kabupaten Muna Barat Berdasarkan Data Potensi Desa Menggunakan Metode *Two Step Cluster*

INFO PENULIS

Riska Nawati Nur
FMIPA Universitas Halu Oleo
riskanawatin@gmail.com

Agusrawati
FMIPA Universitas Halu Oleo
Agusrawati@gmail.com

Gusti Ngurah Adhi Wibawa
FMIPA Universitas Halu Oleo
GustiNgurahAdhiWibawa@gmail.com

INFO ARTIKEL

ISSN: 2808-1307
Vol. 5, No. 3, December 2025
https://jurnal.ardenjaya.com/index.php/ajsh

Saran Penulisan Referensi:

Nur, R. N., Agusrawati., & Wibawa, G. N. A.(2025). Pengelompokan Desa/Kelurahan di Kabupaten Muna Barat Berdasarkan Data Potensi Desa Menggunakan Metode Two Step Cluster. *Arus Jurnal Sosial dan Humaniora, 5 (3),* 5452-5461.

Abstrak

Penelitian ini bertujuan untuk mengelompokkan desa/kelurahan di Kabupaten Muna Barat berdasarkan data Potensi Desa tahun 2023 dan 2024. Pengelompokan dilakukan untuk membantu pemerintah daerah dalam perumusan kebijakan pembangunan yang sesuai dengan karakteristik wilayah.Metode yang digunakan adalah *Two Step Cluster* karena mampu menangani data bertipe campuran (numerik dan kategorik) dalam jumlah besar. Dataset mencakup 86 desa/kelurahan dengan 18 peubah, terdiri atas 10 peubah numerik dan 8 peubah kategorik. Penentuan jumlah *cluster* optimal didasarkan pada nilai *Bayesian Information Criterion (BIC)*, sedangkan kualitas *cluster* dievaluasi dengan Koefisien *Silhouette*.Hasil menunjukkan tiga *cluster* optimal. *Cluster* pertama berisi desa kecil dengan kepadatan tinggi dan fasilitas terbatas. *Cluster* kedua mencakup desa yang luas dengan kepadatan rendah dan minim infrastruktur. *Cluster* ketiga berisi desa yang lebih berkembang, dengan infrastruktur dan layanan publik yang lebih baik. Nilai koefisien *Silhouette* sebesar 0,2 mengindikasikan kualitas *cluster* yang cukup baik.

Kata kunci: Potensi Desa, *Two Step Cluster*, Kabupaten Muna Barat, *BIC*, *Koefisien Silhouette*

Abstract

This study aims to classify villages/urban wards in West Muna Regency based on the Village Potential data from 2023 and 2024. The clustering is intended to assist local governments in formulating development policies aligned with regional characteristics.The method used is the Two-Step Cluster, which is suitable for handling large datasets with mixed data types (numerical and categorical). The dataset comprises 86 villages/urban wards and 18 variables: 10 numerical and 8 categorical. The optimal number of clusters was determined using the Bayesian Information Criterion (BIC), and the clustering quality was assessed using the Silhouette Coefficient.The results show three optimal clusters. The first cluster contains small villages with high population density but limited facilities. The second consists of villages with large land areas and low population density, also facing infrastructure limitations. The third cluster includes more developed villages with better infrastructure and adequate public services. A silhouette coefficient of 0.2 indicates that the clustering quality is fair. These findings are expected to support evidence-based policy-making for regional development.

Keywords: Village Potential, Two-Step Cluster, Clustering, West Muna Regency, BIC, Silhouette Coefficient.

A. Pendahuluan

Desa menurut undang-undang No. 6 Tahun 2014 merupakan kesatuanmasyarakat hukum yang memiliki batas wilayah yang berwenang untuk mengatur sendiri dan mengurus segala urusan pemerintahan, yang berkepentingan untuk masyarakat setempat berdasarkan hak tradisional yang diakui oleh sistem pemerintah negara. Ciri umum desa berdasarkan Pranoto (2001) terbagi menjadi 4 yaitu, dekat dengan pusat wilayah usaha tani, dan kegiatan ekonomi yang paling utama merupakan pertanian, kontrol sosial bersifat informal dan interaksi antar warga desa lebih bersifat personal dalam bentuk tatap muka, serta tingkat homogenitas yang tinggi dan ikatan sosial yang lebih ketat (Pranoto, 2001).

Potensi desa (Podes) adalah kemampuan yang memiliki kemungkinan untuk dikembangkan dalam wilayah otonomi desa. Salah satu cara untuk mengembangkan suatu kabupaten adalah dengan meningkatkan potensi desa. Adanya keberagaman karakteristik di setiap desa maka perlu dilakukan pengelompokan agar memudahkan pemerintah daerah dalam memberikan bantuan dan kebijakan suatu wilayah.

Kabupaten Muna Barat merupakan salah satu daerah otonom di Provinsi Sulawesi Tenggara yang memiliki 11 kecamatan dengan 86 desa/kelurahan. Setiap desa/kelurahan memiliki karakteristik yang berbeda baik dari segi jumlah penduduk, luas wilayah, sarana pendidikan, kesehatan, maupun infrastruktur dasar. Perbedaan ini menimbulkan tantangan bagi pemerintah daerah dalam merumuskan kebijakan pembangunan yang tepat sasaran. Untuk mendukung perencanaan pembangunan berbasis potensi lokal, diperlukan pengelompokan desa yang efektif. Salah satu metode yang digunakan dalam mengelompokkan variabel atau objek adalah metode *cluster*.

Dalam analisis *cluster* ada banyak metode, tergantung jenis data, skala pengukuran, sifat variabel. Pada analisis *cluster* juga banyak dijumpai skala pengukuran variabel tidak sama dan jumlah objek besar serta jumlah *cluster* tidak diketahui. Misalnya pada data potensi desa memiliki skala pengukuran variabel yang berbeda dan mempunyai jumlah data yang besar, maka untuk menangani data potensi desa seperti itu, diperlukan metode *Two Step Cluster. Two Step Cluster* merupakan metode yang cocok untuk data bertipe campuran dan jumlah objek yang besar (Bacher et al., 2004).

Metode *two step cluster* adalah metode yang dirancang untuk mengatasi jumlah objek dengan ukuran yang lebih besar, terutama pada masalah objek dengan skala pengukuran yang berbeda yaitu peubah kontinu dan kategorik*. Two Step Cluster* dipilih karena metode ini secara khusus dirancang untuk data campuran dan dapat menentukan jumlah klaster yang optimal secara otomatis menggunakan kriteria informasi seperti *Bayesian Information Criterion (BIC)*.

Dalam analisis *cluster* penentuan jumlah *cluster* optimal dilakukan dengan *Bayesian Information Criterion (BIC)* dan kualitas hasil klasterisasi dinilai menggunakan Koefisien *Silhouette* (Kaufman & Rousseeuw, 1990).

Beberapa penelitian sebelumnya menunjukkan bahwa metode *Two Step Cluster* mampu menghasilkan klasterisasi yang baik pada data potensi wilayah dan data sosial-ekonomi. Misalnya, Setiawan (2019) berhasil mengelompokkan desa berdasarkan potensi lokal di Kabupaten Konawe, sementara Setiyawati (2020) menggunakannya dalam analisis menu restoran dengan hasil yang cukup memuaskan. Hal ini semakin memperkuat alasan pemilihan metode *Two Step Cluster* dalam penelitian ini.

1. Pengertian Analisis *Cluster*

Analisis *cluster* merupakan analisis untuk menggabungkan objek-objekpengamatan dalam suatu kelompok, dimana objek dalam kelompok memiliki varians homogen namun antar kelompok memiliki varians yang heterogen. Analisis kelompok adalah metode pengelompokan kumpulan data kedalam kelompok yang menampilkan karakteristik serupa (Elvira, 2019).

1) **Metode Pengelompokkan**
   terdapat dua teknik analisis yang digunakan untuk melakukan pengelompokkan data numerik (skala interval dan rasio) yaitu analisis kelompok *hierarki* dan *non-hierarki* serta terdapat pula 1 teknik analisis untuk mengelompokkan data campuran (kategorik dan numerik) yaitu analisis *two step cluster* (Mattjik, 2002).
2) **Metode Hierarki/*Hierarchy Method***
   Metode hierarki (*hierarchical method*) adalah suatu metode pada analisis *cluster* yang

membentuk tingkatan tertentu seperti pada struktur pohon karena 10proses pengklasteran dilakukan secara bertahap atau bertingkat. Hasil Pengklasteran dengan metode *hierarki* dapat disajikan dalam bentuk dendogram. Dendogram adalah representasi visual dari langkah-langkah dalam analisis klaster yang menunjukan bagaimana klaster terbentuk dan nilai koefisien jarak pada setiap langkah (Simamora, 2005).

3) Metode Non hierarki/*Non-hierarchical method*

Pada metode pengelompok kan non berhierarki, peneliti harus menentukan terlebih dahulu jumlah kelompok yang diinginkan. Salah satu contoh dari metode ini adalah metode *K-means.* Analisis kelompok *K-means* menggunakan ukuran jarak *euclidean*. Penentuan pusat kelompok merupakan langkah awal pada metode ini. Langkah selanjutnya adalah menentukan kelompok dari tiap objek, yaitu berdasarkan atas kedekatan ukuran jarak *euclidean* terhadap rata- rata dari masing-masing kelompok (Sari, 2006).

4) Metode *Two Step Cluster*

Metode *two step cluster* merupakan metode yang paling efisien dari segi waktu, hal ini karena metode *two step cluster* didesain untuk data yang berukuran besar. Dalam sebuah penelitian juga dikatakan bahwa analisis *two step cluster* merupakan metode pengklasteran yang dapat memberikan solusi untuk mengelompokkan suatu objek kedalam kelompok-kelompok yang memiliki kemiripan (*homogen*) dengan permasalahan pada skala pengukuran, objek yang diteliti berukuran cukup besar dengan peubah yang bebeda yaitu kategorik dan kontinu sehingga hasil akhir untuk penyelesaian dari metode tersebut dapat diketahui klaster optimal yang terbentuk (Mongi, 2015).

1) Pembentukan *Cluster* Awal (*Pre-Clustering*)

Langkah pertama di metode *two step cluster* ialah pembentukan *cluster* awal *(preclustering)* dilakukan dengan pendekatan secara sekuensial, yaitu setiap objek yang diamati satu persatu bersumber pada ukuran jarak dan setelah itu ditetapkan apakah objek tersebut akan bergabung dalam *cluster* yang sudah terbentuk atau membentuk cluster baru.

Menurut Schiopu (2010), pada pendekatan ini diimplementasikan dengan membentuk *Cluster Feature (CF) Tree. CF Tree* terdiri dari tingkatan cabang (*node*) serta tiap-tiap cabang berisikan beberapa objek/data yang dientrikan. Jika dimisalkan dengan sebuah pohon, sehingga tingkatan cabang tersebut terdiri dari batang pohon, dahan serta daun. Pada *CF Tree* terdapat tingkatan daun yang disebut sebagai daun entri yang berada pada cabang merepresentasikan hasil *sub-cluster* atau anak *cluster*.

Menurut Mongi (2015), prosedur *CF Tree* dapat dilakukan dengan cara memilih satu amatan awal secara random untuk diukur jaraknya masing-masing amatan dengan amatan lain berdasarkan ukuran jarak yang sudah ditetapkan. Apabila besaran jarak letaknya di dalam daerah penerimaan, selanjutnya amatan tersebut akan masuk kedalam anggota anak *cluster*. Apabila besaran jarak letaknya di luar daerah penerimaan, maka amatan tersebut akan masuk kedalam *cluster* yang telah terbentuk atau yang nantinya akan menjadi daun entri baru.

Apabila dalam suatu kasus data/objek ditemui pencilan, maka harus diperiksa apakah *CF Tree* yang terbentuk dapat dimasukkan kedalam *cluster* yang telah terbentuk tanpa harus membentuk lagi *CF Tre*e baru. Untuk mendeteksi pencilan dapat dilakukan dengan perhitungan pada jarak *log-likelihood.* Apabila terdapat jarak terbesar antar cluster yang melebihi nilai titik kritis C, yaitu:

$$C = \log(V) \qquad (2.2)$$

$$V = \Pi_k \mathrm{R}_k \Pi_m L_m \qquad (2.3)$$

dengan:

R_k = range dari peubah kontinu *ke-k*

L_m = banyaknya kategorik untuk peubah kategorik *ke-m.*

Sedangkan pada jarak *Euclidean*, dikatakan data yang memuat pencilan jika jarak terbesar antar cluster melebihi nilai di titik kritis C, rumus C dapat didefinisikan:

$$C = 2\left(\sum_{l=1}^{K^A} \frac{\hat{\sigma}_{kl}^2}{K^A}\right) \wedge \frac{1}{2} \qquad (2.4)$$

dengan:

K^A = banyaknya peubah kontinu

$\hat{\sigma}_{kl}^2$ = ragam dugaan untuk peubah kontinu $ke-l$ dalam cluster $ke-k$.

Pembentukan *CF Tree* terdiri dari dua tahap. Tahapan yang pertama adalah tahap penyisipan (*inserting*) dan tahapan kedua yaitu tahap pembentukan kembali (*rebuilding*). Pada tahapan penyisipan dilakukan secara acak dipilih satu objek kemudian diukur jaraknya ke objek yang lain. Apabila jarak tersebut hasil nilainya kurang dari jarak maksimum, maka objek tersebut akan

dimaksudkan kedalam satu *cluster*. Sedangkan apabila jarak tersebut hasil nilainya melebihi jarak maksimum, maka objek tersebut dikatakan pencilan.

Pada tahapan pembentukan kembali dilakukan berdasarkan objek yang dianggap sebagai pencilan, selanjutnya akan dibentuk suatu *cluster* baru. Apabila *CF Tree* bertambah banyak sehingga melewati batas ukuran maksimum yang telah ditentukan, sehingga batas jarak maksimum perlu ditingkatkan agar dapat memasukkan banyak objek. Peningkatan jarak ini dapat mengakibatkan objek-objek yang berasal dari cluster yang berbeda bergabung menjadi satu *CF*, selanjutnya dihasilkan *CF Tree* dengan ukuran lebih kecil dari sebelumnya. Hasil yang diperoleh dari *CF Tree* diklasterkan dengan analisis *cluster* berhierarki dengan metode penggabungan. Tiap-tiap *cluster* yang terbentuk di tahap pertama akan digabungkan satu persatu berdasarkan ukuran yang telah ditentukan. Prosedur ini akan berakhir sampai seluruh *sub-cluster* mejadi satu *cluster*.

2) Pembentukan *Cluster* Optimal

Menurut Mongi (2015), tahap selanjutnya ialah melakukan tahapan pembentukan *cluster* optimal, dalam menentukan banyaknya jumlah *cluster* dapat dilakukan dengan cara dua tahap, tahap pertama adalah menghitung *Bayesian Information Criterion (BIC) atau Akaike Information Criterion(AIC)* untuk setiap *cluster*.

Bayesian Information Criterion (BIC) dan *Akaike Information Criterion (AIC)* adalah dua kriteria statistik yang digunakan untuk memilih model terbaik dengan mempertimbangkan kecocokan data dan kompleksitas model. Keduanya menghitung likelihood dan memberikan penalti terhadap banyaknya parameter. Namun, *BIC* memberikan penalti yang lebih besar, sehingga lebih selektif dalam memilih model. Dalam analisis *Two Step Cluster***,** metode *BIC* lebih umum digunakan karena memberikan hasil yang cenderung lebih stabil dan sederhana, serta merupakan default dalam perangkat lunak seperti SPSS. Oleh karena itu, penelitian ini menggunakan *BIC* sebagai kriteria utama dalam menentukan jumlah klaster optimal.

Rumus *BIC* dan *AIC* untuk *cluster J* dapat didefinisikan sebagai berikut:

$$m_j = J\left\{2K^A + \sum_{k=1}^{K^B}(L_k - 1)\right\} \quad (2.5)$$

$$\xi_j = -N\left(\sum_{k=1}^{K^A}\frac{1}{2}\log\left(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2\right)\sum_{k=1}^{K^B}\sum_{k=1}^{L_k}\frac{N_{jkl}}{N_j}\log\left(\frac{N_{jkl}}{N_j}\right)\right) \quad (2.6)$$

$$BIC(J) = -2\sum_{j=1}^{J}\xi_j + m_j\log(N) \quad (2.7)$$

$$AIC(J) = -2\sum_{j=1}^{J}\xi_j + 2m_j \quad (2.8)$$

dengan:

N = jumlah data

N_j = jumlah data di dalam *cluster j*

N_{jkl} = jumlah data di *cluster j* untuk peubah kategorik *ke-k* dengan *ke-l*

$\hat{\sigma}_k^2$ = ragam dugaan untuk peubah kontinu *ke-k* untuk keseluruhan data

$\hat{\sigma}_{jk}^2$ = ragam dugaan untuk peubah kontinu *ke-k* dalam *cluster j*

K^A = jumlah peubah kontinu

K^B = jumlah peubah kategorik

L_k = jumlah kategorik untuk peubah kategorik *ke-k.*

Selanjutnya hasil pada perhitungan tersebut akan digunakan untuk menduga jumlah *cluster*. Tahap kedua adalah mencari jarak terbesar antara dua *cluster* yang saling berdeketan pada tiap-tiap tahapan dalam pengklasteran. Solusi banyaknya *cluster* terbaik adalah *BIC* yang memiliki nilai terkecil, namun ada beberapa masalah di dalam pengklasteran dimana nilai *BIC* akan terus bertambah jika jumlah *cluster* semakin meningkat. Sehingga dalam kasus tersebut, rasio perubahan *BIC (ratio BIC changes)* dan rasio perubahan jarak (*ratio of distance measure changes*) yang mana nilai tersebut untuk mengidentifikasi solusi dari banyaknya *cluster* optimal. Solusi untuk banyaknya *cluster* optimal akan memiliki rasio perubahan *BIC* dan rasio perubahan jarak yang besar.

Perbandingan antar jarak untuk *k cluster* digunakan untuk mengetahui banyaknya cluster yang terbentuk, dengan rumus perbandingan sebagai berikut :

$$R(k) = \frac{d_{k-1}}{d_k} \quad (2.9)$$

$$d_k = l_{k-1} - l_k \quad (2.10)$$

$$l_v = \frac{r_v \log n - BIC_v}{2} \quad (2.11)$$

$$l_v = \frac{(2r_v - AIC_v)}{2} \quad (2.12)$$

$$v = k, k-1 \quad (2.13)$$

dengan:

d_{k-1} = jarak jika *k* cluster digabungkan dengan *J – 1 cluster*

$R(k)$ = rasio perubahan jarak

(Setiawan dan Pratiwi, 2019).

2. Ukuran Jarak

Ukuran jarak merupakan ukuran kemiripan atau ketakmiripan yang digunakan dalam metode *cluster* adalah jarak objek dan jarak antar *cluster*.Penentuan jarak dapat menggunakan beberapa metode pengukuran seperti jarak *Euclidean* atau *log-likehood.* Akan tetapi, jarak *Euclidean* hanya dapat menangani data kontinu. Sedangkan pada penelitian ini memiliki data campuran sehingga penentuan ukuran jarak menggunakan jarak *log-likelihood* (Şchiopu, 2010). Ukuran jarak diperlukan untuk setiap pasang objek yang akan dikelompokkan. Ada beberapa metode pengukuran jarak antar dua objek, yaitu:

1) Jarak *Euclidean*

Jarak Euclidean merupakan jarak yang paling umum dan sering digunakan dalam analisis *cluster* apabila peubahnya berskala kontinu. Jarak Euclidean harus memenuhi asumsi jika peubah-peubah yang diamati tidak berkolerasi dan antar peubah memiliki antar satuan yang sama. Pada metode ini, pengukuran jarak dilakukan dengan menghitung akar kuadrat dari penjumlahan kuadrat selisih dari nilai masing-masing peubah. Jarak Euclidean dapat didefinisikan sebagai berikut:

$$d_{i,j} = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2} \quad (2.14)$$

dengan:

$d_{i,j}$ = Jarak antara objek *i* dengan objek *j*

x_{ik} = Nilai objek *i* pada peubah ke-*k*

x_{jk} = Nilai objek *j* pada peubah ke-*k*

p = Banyaknya peubah yang diamati

(Mongi, 2015).

2) Jarak *Log-likelihood*

Jarak *log-likelihood* adalah jarak yang digunakan untuk peubah skala kontinu dan kategorik. Ukuran jarak *log-likelihood* digunakan untuk data bertipe campuranyang kekar terhadap pelanggaran asumsi kebebasan dan distribusi data (Batcher *et al*, 2004). Dapat digunakan untuk peubah kontinu maupun kategorik. Berikut merupakan rumus jarak *log-likelihood* (Li & Sun, 2018). Jarak antar klaster j dan s dapat dirumuskan sebagai berikut:

$$d(j,s) = \xi_j + \xi_s + \xi_{(j,s)} \quad (2.15)$$

dengan:

$$\xi_j = -N\left(\sum_{k=1}^{K^A} \frac{1}{2}\log\left(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2\right) - \sum_{k=1}^{K^B}\sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j}\log\left(\frac{N_{jkl}}{N_j}\right)\right)$$

$$\xi_s = -N\left(\sum_{k=1}^{K^A} \frac{1}{2}\log\left(\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2\right) - \sum_{k=1}^{K^B}\sum_{l=1}^{L_k} \frac{N_{skl}}{N_j}\log\left(\frac{N_{skl}}{N_j}\right)\right)$$

$$\xi_{js} = -N\left(\sum_{k=1}^{K^A} \frac{1}{2}\log\left(\hat{\sigma}_k^2 + \hat{\sigma}_{(js)k}^2\right) - \sum_{k=1}^{K^B}\sum_{l=1}^{L_k} \frac{N_{(js)kl}}{N_j}\log\left(\frac{N_{(js)kl}}{N_j}\right)\right)$$

dengan:

N = Banyaknya objek

N_j = Jumlah objek didalam *cluster j*

N_{jkl} = Jumlah objek di *cluster j* untuk peubah kategorik ke-*k* dengan kategori ke-*l*

$\hat{\sigma}_k^2$ = Ragam dugaan untuk peubah kontinu ke-*k* untuk keseluruhan objek

$\hat{\sigma}_{jk}^2$ = Ragam dugaan untuk peubah kontinu ke-*k* untuk keseluruhan objek dalam cluster *j*

K^B = Banyaknya peubah kontinu

K^A = Banyaknya peubah kategorik

L_k = Banyaknya kategori untuk peubah kategorik ke-*k*

(Mongi, 2015)

3) Koefisien *Silhoutte*

Koefisien *Silhoutte* dapat digunakan untuk memeriksa objek ke dalamsetiap cluster. Nilai koefisien ini berkisar antara -1 hingga 1. Nilai *Silhouette Coefficient* yang negatif berarti objek yang diukur lebih terkait dengan objek di cluster lain. Semakin rendah nilai SC

semakin jauh jarak antara objek dengan cluster. Fokus utama pada tahap ini adalah tingkat kepentingan peubah terhadap klaster yang terbentuk. Tingkat kepentingan peubah diukur dengan skala 0 sampai 1, dengan 0 berarti sangat tidak penting dan 1 sangat penting. Koefisien *Silhoutte* Merupakan teknik grafik yang digunakan untuk mengukur kualitas klaster. Nilai rata-rata Koefisien *Silhoutte* berada antara -1 sampai 1.

Kriteria subjektif pengukuran pengelompokan berdasarkan nilai *SilhoutteCoefficient* (SC), dapat dilihat pada tabel berikut untuk rentang nilai dan kriteria (Kauffman and Roesseeuw, 1990).

Tabel 2.1 Kriteria pengukuran koefisien *Silhoutte*

Nilai SC	Kriteria
-1— 0,2	Kulitas Buruk
0,2 – 0,5	Kualitas cukup
0,5 – 1	Kulitas Baik

Untuk rata-rata koefisien *Silhouette* dapat digunakan untuk mengukur kualitas klasterisasi. Nilainya juga berkisar antara -1 hingga 1. Jika nilai koefisien *Silhouette* antara -1 hingga 0,2 maka kualitasnya diklasifikasikan buruk, 0,2 sampai 0,5 berada pada kondisi cukup, dan 0,5 sampai 1 berada padakondisi baik (Supandi *et al*., 2021).

Rumus koefisien *Silhouette* dapat dilihat sebagai berikut:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (2.16)$$

$$\overline{\boldsymbol{S}} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{S_i} \quad (2.17)$$

dengan**:**

s_i = Koefisien *Silhoutte* untuk setiap objek ke *i*

b_i = Rata-rata jarak minimum antar objek ke-*i* pada klaster yang berbeda

a_i = Rata-rata jarak minimum antar objek ke-*i* pada klaster yang sama

$\bar{S}$ = Nilai rata-rata koefiseien *Silhoutte*

N = Total amatan

Prosedur pengelompokan mempunyai dua tahapan pengelompokan awal objek ke dalam *subklaster-subklaster* kecil dan tahap pengelompokan optimal (Mongi, 2015).

4) Keterangan Umum Desa

di desa merupakan pembangunan yang dilaksanakan secara menyeluruh dan terpadu dengan kewajiban yang serasi antara pemerintah dan masyarakat, dimana pemerintah wajib memberikan bimbingan, pengarahan, bantuan, dan fasilitas yang diperlukan. Sedangkan masyarakat memberikan partisipasinya dalam membangun potensi gotong-royong masyarakat pada setiap pembagunan yang diinginkan untuk meningkatkan pendapatan dan kesejahteraanmasyarakat di pedesaan (Maryani & Waluya, 2008).

5) Potensi desa

Potensi desa (Podes) adalah kemampuan yang memiliki kemungkinan untuk dikembangkan dalam wilayah otonomi desa. Datapotensi merupakan satu- satunya data yang berurusan dengan wilayah atau tata ruang dengan basis desa atau kelurahan (BPS, 2020). Selain itu, data podes juga merupakan sumber data kewilayahan yang muatannya beragam dan memberi gambaran tentang situasi pembangunan suatu wilayah (*regional*). Pengembangan potensi desa bertujuan untuk mendorong kemandirian masyarakat desa/kelurahan melalui pengembangan potensi unggulan dan penguatan kelembagaan serta pemberdayaan masyarakat (Abdurokhman, 2015).

B. Metodologi

Data yang digunakan adalah data sekunder dari BPS Kabupaten Muna Barat tahun 2023 dan 2024. Jumlah unit analisis meliputi 86 desa/kelurahan dan 18 peubah (10 numerik dan 8 kategorik).

Tabel 1. Peubah Penelitian

Peubah	Deskripsi peubah	Tipe

X1	Status Desa (1 = Desa, 2 = Kelurahan)	Kategorik
X2	Kondisi Topografi (1 = Dataran, 2 = Lereng, 3 = Lembah)	Kategorik
X3	Jenis Permukaan Jalan Terluas (1 = Aspal, 2 = Diperkeras, 3 = Tanah, 4 = Kayu)	Kategorik
X4	Jenis Prasarana Transportasi (1 = Darat, 2 = Darat/Air, 3 = Air)	Kategorik
X5	Keberadaan Angkutan Umum (1 = Ada, dengan trayek, 2 = Ada tanpa trayek, 3 = Tidak ada)	Kategorik
X6	Kekuatan Sinyal Telepon Seluler (1 = Sangat kuat, 2 = kuat, 3 = lemah)	Kategorik
X7	Jenis Sinyal Internet Telepon Seluler (1 = 4G/LTE, 2 = 3G/H/H+/EVDO, 3 = 2,5G/E/GPRS)	Kategorik
X8	Menara Telepon Seluler (1 = Ada, 2 = Tidak ada)	Kategorik
X9	Luas Daerah (Km^2)	Numerik
X10	Jumlah penduduk (jiwa)	Numerik
X11	Kepadatan penduduk per km^2	Numerik
X12	Ketinggian (DPL)	Numerik
X13	Jarak ke ibukota kecamatan (km)	Numerik
X14	Jarak ke ibukota kabupaten (km)	Numerik
X15	Jumlah fasilitas pendidikan (sekolah)	Numerik
X16	Jumlah pengguna listrik (rumah tangga)	Numerik
X17	Jumlah sarana perdagangan (pasar)	Numerik
X18	Jumlah tempat ibadah (gereja, masjid)	Numerik

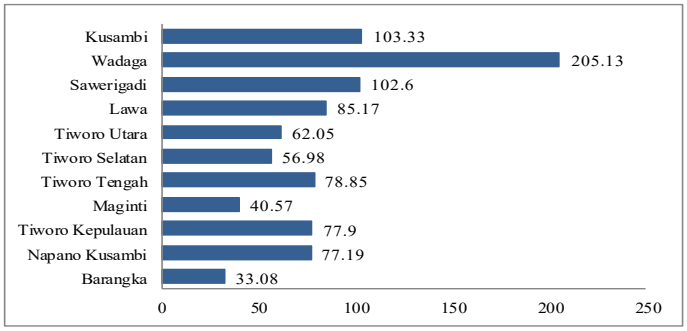
Tahapan analisis:

1. Deskripsi peubah penelitian
2. *Pre-clustering* menggunakan SPSS dengan ukuran jarak *log-likelihood*.
3. Penentuan jumlah klaster optimal berdasarkan nilai *BIC*.
4. Evaluasi kualitas klaster menggunakan *Koefisien Silhouette*.

C. Hasil dan Pembahasan

1. Deskripsi Umum

Kabupaten Muna Barat memiliki 11 kecamatan dengan 86 desa/kelurahan. Mayoritas berada di dataran dan menggunakan moda transportasi darat. Permukaan jalan dominan berupa aspal, dan sebagian besar wilayah telah memiliki infrastruktur dasar. Dengan luas wilayah sebesar 818,70 km^2 dengan rincian luas kecamatan disajikan pada Gambar 1 berikut.



Gambar 1 Luas Kecamatan di Kabupaten Muna Barat

Dari gambar 1 diketahui bahwa kecamatan Wadaga merupakan kecamatan yang terluas dengan luas daerah sebesar 205,13 km^2 sedangkan kecamatan barangka merupakan kecamatan terkecil dengan luas daerah sebesar 33,08 km^2.

2. Pengelompokan

Nilai *BIC* dihitung berdasarkan nilai *log-likelihood* dari model dan jumlah parameter yang digunakan. Model dengan nilai *BIC* terkecil dianggap sebagai model terbaik karena memberikan keseimbangan antara kesesuaian model dengan data (*goodness of fit*) dan kompleksitas model yang seminimal mungkin.

Tabel 2 Nilai BIC dari banyaknya *cluster* yang terbentuk

Jumlah Cluster	Bayesian Criterion (BIC)	Perubahan BIC	Rasio Perubahan BIC	Rasio Perubahan Ukuran Jarak

1	1674.092			
2	1576.098	-97.994	1.000	1.451
3	1557.047	-19.051	.194	1.273
4	1575.546	18.499	-.189	1.977
5	1661.958	86.412	-.882	1.140
6	1756.922	94.963	-.969	1.044
7	1854.436	97.515	-.995	1.138
8	1959.022	104.586	-1.067	1.048
9	2065.956	106.933	-1.091	1.215
10	2181.564	115.608	-1.180	1.042
11	2298.788	117.224	-1.196	1.060
12	2418.193	119.405	-1.218	1.316
13	2546.356	128.164	-1.308	1.050
14	2675.832	129.476	-1.321	1.071
15	2807.062	131.230	-1.339	1.175

Berdasarkan Tabel 2, *nilai Bayesian Information Criterion (BIC)* mengalami penurunan signifikan dari *cluster* 1 ke *cluster* 2, dengan perubahan sebesar -97,994 dan rasio perubahan ukuran jarak sebesar 1,451. Penurunan BIC kembali terjadi hingga jumlah *cluster* ke-3 dengan nilai minimum sebesar 1557,047. Setelah *cluster* ke-3, nilai *BIC* justru meningkat dan rasio perubahan *BIC* menjadi lebih kecil dari 1, menunjukkan bahwa penambahan *cluster* tidak memberikan peningkatan model yang signifikan

Berdasarkan perhitungan *BIC*, ditemukan bahwa jumlah klaster optimal adalah 3, dengan rincian:

Tabel 3. *Distribusi Desa/Kelurahan Berdasarkan Klaster*

Jumlah Cluster	Desa	Kelurahan	Total (N)	Persentase
1	11	0	11	12.8%
2	19	0	19	22.1%
3	51	5	56	65.1%
Total	**81**	**5**	**86**	**100%**

Nilai koefisien Silhouette sebesar 0,2 menandakan kualitas klaster yang cukup baik.

D3. Interpretasi *Cluster*

- ***Cluster* 1:** Desa berukuran kecil, kepadatan tinggi, dan fasilitas terbatas.
- ***Cluster* 2:** Desa luas dengan kepadatan rendah dan minim infrastruktur.
- ***Cluster* 3:** Desa berkembang dengan infrastruktur memadai dan pelayanan publik yang lengkap.

Karakteristik numerik dan kategorik pada tiap *cluster* memperkuat pemahaman mengenai pola perkembangan desa di wilayah ini.

Tabel 4. Karakteristik dari peubah numerik yang berpengaruh

Cluster	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18
1	4.1173	809.18	908.48	2.00	19.52	39.51	2.55	216.82	0.45	1.00
2	12.7637	795.00	107.02	29.00	11.06	10.88	2.84	192.74	0.21	1.68
3	10.8030	1123.46	155.63	34.87	3.44	16.48	3.27	303.59	1.18	2.68

Tabel 4 menyajikan karakteristik numerik seperti luas wilayah, jumlah penduduk, fasilitas publik, dan rumah tangga pengguna listrik pada masing-masing *cluster*. *Cluster* 1 memiliki desa-desa kecil dengan kepadatan penduduk tinggi namun jumlah fasilitas terbatas. *Cluster* 2 memiliki wilayah yang luas dan fasilitas minim, sedangkan *cluster* 3 terdiri dari desa-desa yang lebih berkembang dengan fasilitas publik, perdagangan, dan pemukiman yang lebih lengkap.

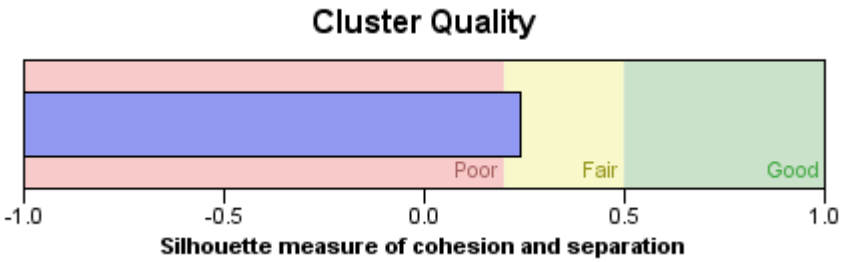
Tabel 5. Karakteristik dari peubah kategorik yang berpengaruh

Peubah	Kategori	Klaster 1 (n=11)	%	Klaster 2 (n=19)	%	Klaster 3 (n=56)	%
Status Desa (X1)	Desa	11	100%	19	100%	51	91.1%
	Kelurahan	0	0%	0	0%	5	8.9%

Kondisi Topografi (X2)	Dataran	11	100%	6	31.6%	23	41.1%
	Lereng	0	0%	12	63.2%	28	50.0%
	Lembah	0	0%	1	5.3%	5	8.9%
Permukaan Jalan (X3)	Aspal	11	100%	0	0%	55	98.2%
	Diperkeras	0	0%	6	31.6%	0	0%
	Tanah	0	0%	6	31.6%	1	1.8%
	Kayu	0	0%	7	36.8%	0	0%
Jenis Transportasi (X4)	Darat	5	45.5%	18	94.7%	49	87.5%
	Darat/Air	5	45.5%	0	0%	6	10.7%
	Air	1	9.1%	1	5.3%	1	1.8%
Keberadaan Angkutan (X5)	Ada dengan trayek	0	0%	3	15.8%	44	78.6%
	Ada tanpa trayek	5	45.5%	3	15.8%	8	14.3%
	Tidak ada	6	54.5%	13	68.4%	4	7.1%
Kekuatan Sinyal (X6)	Sangat kuat	4	36.4%	0	0%	10	17.9%
	Kuat	0	0%	5	26.3%	35	62.5%
	Lemah	7	63.6%	14	73.7%	11	19.6%
Jenis Sinyal (X7)	4G/LTE	11	100%	0	0%	38	67.9%
	3G/H+/EV DO	0	0%	6	31.6%	18	32.1%
	2.5G/GPRS	0	0%	13	68.4%	0	0%
Menara Telekomunikasi (X8)	Ada	1	9.1%	1	5.3%	12	21.4%
	Tidak ada	10	90.9%	18	94.7%	44	78.6%

Tabel 5 menggambarkan karakteristik kategorik seperti status desa, kondisi jalan, sinyal komunikasi, dan keberadaan fasilitas umum lainnya. Desa pada *cluster* 1 umumnya berada di dataran dengan jalan beraspal dan sinyal 4G, tetapi tidak memiliki menara telekomunikasi dan akses angkutan umum. *Cluster* 2 didominasi desa dengan kondisi geografis dan infrastruktur terbatas. Sementara *cluster* 3 menunjukkan variasi yang lebih tinggi dengan dukungan fasilitas dan akses teknologi yang lebih baik.

4. Kualitas *cluster* yang terbentuk



Gambar 2. Kualitas *cluster* yang terbentuk

Berdasarkan gambar di atas, didapatkan bahwa koefisien *silhouette* bernilai 0,2 yang artinya kualitas clusternya cukup baik. Meskipun tidak ideal (>0,5), nilai ini masih dapat diterima untuk analisis eksploratif, mengingat data mencakup variabel campuran dan jumlah objek cukup besar. Untuk mengetahui peubah yang digunakan pada uji selanjutnya digunakan peubah yang memiliki tingkat kepentingan yang cukup tinggi.

D. Kesimpulan

Metode *Two Step Cluster* berhasil mengelompokkan desa/kelurahan di Kabupaten Muna Barat menjadi tiga *cluster* yang merepresentasikan karakteristik sosial dan geografis wilayah. Tiga *cluster* desa yang terbentuk mencerminkan keberagaman karakteristik wilayah di Kabupaten Muna Barat.

Dari hasil analisis klaster menggunakan metode *Two-Step Cluster* terhadap 86 desa/kelurahan, diperoleh tiga klaster optimal yang menggambarkan karakteristik wilayah berdasarkan kombinasi variabel kategorik dan numerik. Proses clustering menunjukkan bahwa sebanyak 65,1% unit wilayah tergolong dalam *Cluster* 3, 22,1% dalam *Cluster* 2, dan 12,8% dalam *Cluster* 1.

Kualitas hasil klaster berdasarkan nilai silhouette sebesar 0,2 berada pada kategori cukup, yang menunjukkan bahwa pemisahan antar klaster belum sepenuhnya optimal. Oleh karena itu, meskipun hasil klaster memberikan gambaran umum yang berguna, interpretasinya perlu dilakukan secara hati-hati dan dapat dilengkapi dengan pendekatan analisis tambahan untuk mendukung rekomendasi kebijakan pembangunan wilayah yang lebih akurat.

Hasil klasterisasi ini diharapkan menjadi dasar awal bagi pemerintah daerah dalam menyusun kebijakan pembangunan berbasis klaster wilayah.

Saran

Pada penelitian ini hanya mengkaji mengenai metode *two step cluster*. Oleh karena itu peneliti juga ingin melakukan *clustering* dengan metode lainnya mengingat cakupan metode analisis *cluster* yang cukup banyak serta dapat dikembangkan dengan mengaplikasikan pada bidang ilmu lainnya.

E. Referensi

Abdurokhman, M. (2015). Analisis Potensi Desa dalam Pengelompokan Wilayah di Kabupaten Tasikmalaya. *Jurnal Geografi*, 12(2), 45–53.

Bacher, J., Wenzig, K., & Vogler, M. (2004). *SPSS TwoStep Cluster – A First Evaluation*. Universität Erlangen-Nürnberg.

Badan Pusat Statistik. (2020). *Potensi Desa 2020*. Jakarta: BPS RI.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.

Maryani, E., & Waluya, S. B. (2008). Analisis Kelayakan Potensi Desa untuk Pembangunan Berkelanjutan. *Jurnal Pembangunan Wilayah*, 4(1), 11–20.

Mattjik, A. A. (2002). *Perancangan Percobaan*. IPB Press.

Mongi, V. M. (2015). Pemanfaatan Data Potensi Desa dalam Perencanaan Wilayah. *Jurnal Perencanaan Pembangunan*, 7(2), 25–34.

Pranoto, B. (2001). *Statistik Multivariat*. Gadjah Mada University Press.

Sari, R. N. (2006). Penerapan Analisis Cluster untuk Pengelompokan Daerah. *Jurnal Statistika*, 8(1), 55–63.

Setiawan, A., & Pratiwi, N. (2015). Analisis Cluster Menggunakan Metode Two Step Cluster. *Jurnal Statistika Universitas Diponegoro*, 3(1), 12–18.

Setiyawati, R. (2016). Evaluasi Kualitas Klaster dengan Koefisien Silhouette. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika*, 223–230.

Supandi, S., Yuniarto, Y., & Putri, R. (2018). Penerapan Metode Two Step Cluster dalam Pengelompokan Data Potensi Desa. *Jurnal Matematika dan Aplikasi*, 17(1), 33–40.