
QUICK ROBUST CLUSTERING USING LINKS (QROCK) UNTUK PENGELOMPOKAN DESA KABUPATEN BANJAR

Muhammad Rizki Shofari^{1*}, Oni Soesanto², Dwi Kartini³

¹ Program Studi Statistika Fakultas MIPA Universitas Lambung Mangkurat

² Program Studi Matematika Fakultas MIPA Universitas Lambung Mangkurat

e-mail: muhammadrizkishofari@gmail.com

Abstract

Accurate village profile planning needs to be done with mapping based on its characteristics. Village grouping based on the characteristics of village facilities and potential or its characteristics based on the Building Village Index indicator can help determine priorities in village development. In this study, mixed data was used, with numerical data grouping using Hierarchical Agglomerative Nesting (AGNES) algorithm and categorical data with Quick Robust Clustering Using Links (QROCK). The resulting clusters are then combined using the QROCK Ensemble algorithm (algCEBMDC). The data is sourced from the 2021 Village Potential Data Collection (PODES) by the Central Statistics Agency in 277 villages in Banjar Regency, including 18 numerical variables and 29 categorical variables. The results of the study obtained optimal clusters based on the ratio of within-group standard deviation (SW) to between-group standard deviation (SB) resulting in a ratio of 4.82.10⁻⁹ with a threshold of 0.4 to 0.9 resulting in 6 clusters. The best cluster results are cluster 4 (4 villages) and cluster 3 (14 villages), then cluster 2 (villages) and cluster 1 (186 villages), and clusters that need development priorities are cluster 5 (2 villages) and cluster 6 (1 village) which are outliers based on processing results.

Keywords: Village Grouping, Cluster, AGNES Algorithm, Quick Robust Clustering Using Links (QROCK), AlgCEBMDC

1. PENDAHULUAN

Pembangunan desa menjadi fokus utama pemerintah dalam pengesahan undang-Undang Desa No. 6 Tahun 2014 tentang Rencana Pembangunan Jangka Menengah (RPJM Desa) yang memiliki memiliki tujuan untuk meningkatkan kesejahteraan masyarakat dan penanggulangan kemiskinan [10]. Badan Pusat Statistik (BPS) mendata potensi desa di seluruh Indonesia untuk menyusun rencana pembangunan yang sesuai dengan karakteristik setiap desa. Data yang dihasilkan dari pendataan ini digunakan untuk membentuk Indeks Desa Membangun (IDM) yang kini menjadi pedoman utama bagi pemerintah dalam menangani Desa Tertinggal.

Kabupaten Banjar sebagai salah satu wilayah terluas di Provinsi Kalimantan Selatan perlu bekerja lebih keras untuk menyediakan infrastruktur, aksesibilitas, dan pelayanan publik yang merata di seluruh desa dengan jumlah desa yang sangat banyak yaitu 277 desa [1]. Untuk menetapkan perencanaan yang tepat maka dilakukan Analisis Pengelompokan (Cluster Analysis) yang bertujuan untuk membagi objek-objek diamati ke dalam cluster menurut masing-masing karakteristiknya.

IDM tersusun atas indikator data bertipe numerik & kategorik. Metode ensemble dapat digunakan untuk data campuran melalui teknik penggabungan hasil pengelompokan dari pelbagai algoritma [8]. Pengelompokan data numerik menggunakan metode Algoritme Hierarchical Agglomerative Nesting (AGNES), dan pengelompokan data kategorik menggunakan metode Quick Robust Clustering Using Links (QROCK). Kemudian masing-masing hasil pengelompokan digabungkan (Ensemble) dan

diproses menggunakan algoritme pengelompokan data kategorik (Ensemble QROCK) guna mendapatkan hasil akhir.

2. TINJAUAN PUSTAKA

2.1. Data

Data merupakan suatu fakta dan informasi yang diperoleh melalui pengamatan, pengukuran, dan penelitian pada variabel-variabel tertentu. Data statistik digunakan untuk menyajikan, menggambarkan, dan menyimpulkan informasi tentang populasi atau sampel yang relevan [16]. Data dibedakan berdasarkan sifatnya yaitu data numerik atau kuantitatif dan kategorik atau kualitatif. Adapun data numerik berupa data terdiri dari angka atau bilangan yang dapat dikenali sebagai kuantitas atau besaran [18]. Data kategorik merupakan data simbolik berupa tipe data yang tidak bisa dihitung secara matematis [3].

2.2. Statistika Deskriptif

Statistika deskriptif ialah suatu proses mengumpulkan dan menyajikan data sehingga menghasilkan informasi bermanfaat. [14]. Statistika deskriptif memberikan gambaran karakteristik data.

2.3. Analisis Kelompok

Analisis kelompok merupakan teknik pengelompokan sekumpulan n objek kedalam m *cluster* yang memiliki kemiripan dengan objek lainnya dan memiliki perbedaan dari objek yang berada pada semua *cluster* lainnya. Hasil analisis kelompok mengharapkan kesamaan yang tinggi antara anggota dan perbedaan yang tinggi antara anggota kelompok [6]. Pengelompokan didasarkan pada kemiripan atau ketakmiripan objek yang diamati; ukuran kemiripan dan ketakmiripan biasanya diukur dengan jarak [15].

2.3.1. Pengukuran Jarak Numerik

pengukuran ketakmiripan jarak variabel numerik yang biasanya digunakan adalah jarak *euclidean* dua objek adalah [18]:

$$d(i, j) = \sqrt{\sum_{n=1}^k (X_i - X_j)^2} \quad (1)$$

$d(i, j)$ jarak objek ke - i dengan objek ke - j
 k banyak *cluster*
 X_i data subjek ke - i peubah ke - n
 X_j data subjek ke - j peubah ke - n

2.3.2. Pengukuran Jarak Kategorik

Pengukuran kemiripan jarak variabel kategorik yang digunakan adalah *weighted similarity measure*. Jarak *weighted similarity measure* antar dua objek adalah [4]:

$$sim(X_i X_j) = \frac{|X_i \cap X_j|}{|X_i \cap X_j| + 2 \sum_{k \in X_i \cap X_j} \frac{1}{|D_k|}} \quad (2)$$

$|X_i \cap X_j|$ Banyak kategori sama antar X_i dengan X_j
 $|D_k|$ Selisih perbedaan tingkat kategori antara X_i dan X_j

2.4. Pengelompokan Data Numerik

Metode hierarki dan non-hierarki digunakan untuk pengelompokan data numerik. Jika banyak cluster yang akan dibentuk tidak diketahui, maka menggunakan metode hierark. Teknik pengelompokan *cluster* hirarki terdapat dua, yaitu *divisive* dan *agglomerative* [11]. Pengelompokan hierarki *agglomerative* merupakan metode clustering yang dimulai dengan setiap titik data dianggap sebagai satu kluster terpisah dan secara berangsur-angsur menggabungkan kluster yang paling dekat satu sama lain. Beberapa teknik pengelompokan adalah sebagai berikut [9]:

a. *Single Linkage*

Proses menggabungkan dua objek menggunakan jarak terdekat. Adapun ukuran jaraknya :

$$d_{(ij)k} = \min (d_{ik}, d_{jk}) \quad (3)$$

d_{ik} jarak terdekat dari *cluster* i dan k
 d_{jk} jarak terdekat dari *cluster* j dan k

b. *Complete Linkage*

Dalam proses ini, bergantung pada jarak terbesar atau jarak terjauh untuk menggabungkan dua objek. Prosedur ini berlawanan dengan *Single Linkage*. Ukuran jaraknya adalah :

$$d_{(ij)k} = \max (d_{ik}, d_{jk}) \quad (4)$$

d_{ik} jarak terjauh dari *cluster* i dan k
 d_{jk} terjauh dari *cluster* j dan k

c. *Average Linkage*

Proses menggunakan ukuran jarak rata-rata antara semua objek. Ukuran jaraknya adalah :

$$d_{(ij)k} = \frac{\sum_{i=1}^n \sum_{k=1}^n d_{ik}}{n_{(ij)} n_k} \quad (5)$$

d_{ik} Jarak objek i pada *cluster* (uv) dengan objek j pada *cluster* w
 $n_{(ij)}$ Jumlah objek pada *cluster* (ij)
 n_k Jumlah objek pada *cluster* k

2.5. Pengelompokan Data Kategorik

ROCK (Robust Clustering with Links) adalah alternatif metode pengelompokan data kategorik [5]. Metode ROCK, di sisi lain, dianggap kurang efektif dalam menangani outlier. Sehingga Dutta pada tahun 2005 memperkenalkan penyempurnaan dari metode ROCK yaitu QROCK (*Quick Robust Clustering with links*).

Untuk menggabungkan cluster-clusternya, QROCK menggunakan tipe data abstrak yang dikenal sebagai MFSET. Tahap MFSET yaitu [4]:

- Merge* : combining components A and B
Find (x) : search for components whose members are x
Initial (x) : forms the components that make up element x

algoritma QROCK dinilai dapat mendeteksi adanya *outlier* secara efektif. Langkah-langkah pengelompokan data kategorik menggunakan metode QROCK yaitu [4]:

- 1) Menginisialisasi nilai θ atau besar nilai similaritas awal yang berkisar dari 0 hingga 1.

- 2) Menghitung kemiripan (similaritas) terbobot menggunakan persamaan:

$$sim(X_i X_j) = \frac{|X_i \cap X_j|}{|X_i \cap X_j| + 2 \sum_{k \in X_i \cap X_j} \frac{1}{|D_k|}} \quad (6)$$

- $|X_i \cap X_j|$ Banyak kategori sama antar X_i dan X_j
 $|D_k|$ Selisih tingkat perbedaan kategori antara X_i dan X_j

- 3) Mencari *neighbors* dari masing - masing objek.

X_i dan X_j bertetangga jika $sim(X_i X_j) \geq \theta$. *Neighbors* dari masing - masing objek :

$$neighbors [X_1] = [X_i, X_1]$$

$$neighbors [X_2] = [X_i, X_2]$$

$$neighbors [X_3] = [X_i, X_3]$$

⋮

$$neighbors [X_j] = [X_i, X_j]$$

Setiap objek X_i akan memiliki daftar tetangga (*neighbors* [X_i]) yang terdiri dari objek-objek yang memenuhi syarat kemiripan di atas ambang batas.

- 4) Melakukan pengelompokan.

- i. Inisialisasi (i) pada tiap - tiap objek.

- ii. Sampai dengan n untuk setiap $i, i = 1$

Ambil nilai x pada *neighbours* [X_i] = [X_i, X_j] $\Rightarrow x = X_i$

Ambil nilai $y = x$ dengan *neighbours* [X_i] = [X_i, X_j] $\Rightarrow y = X_j$

$A = find(x) = find(T_i) = X_i$

$B = find(y) = find(T_j) = X_j$

Jika $A \neq B \Rightarrow merge(A, B) = X_i, X_j$, sehingga terbentuk sekumpulan komponen $\{X_i, X_j\}, \dots, \{X_j\}$.

Jika $A = B \Rightarrow$ maka tidak digabungkan

- iii. Pengulangan langkah (ii) hingga $i = n$; iterasi selesai.

2.6. Cluster Ensemble

Cluster ensemble adalah teknik yang menggabungkan berbagai algoritma untuk mendapatkan hasil akhir dari menyatukan hasil beberapa masing - masing metode pengelompokan [19]. Langkah-langkah analisis data campuran metode *cluster Ensemble* yang disebut algoritme CEBMDC memiliki tahapan sebagai berikut [8] :

- 1) Data campuran dibagi menjadi dua bagian data yaitu data numerik dan data kategorik.
- 2) Melakukan pengelompokan masing – masing algoritme pengelompokan untuk variabel numerik, dan algoritme pengelompokan untuk variabel kategorik.
- 3) Proses ensemble yaitu penggabungan hasil metode pengelompokan numerik dan pengelompokan kategorik.
- 4) Mendapatkan cluster akhir menggunakan algoritme pengelompokan data kategorik dari hasil penggabungan data campuran.

2.7. Kinerja Hasil Pengelompokan Numerik

Salah satu cara untuk memastikan validitas suatu pengelompokan adalah dengan mengukur kinerja hasilnya. Nilai maksimum *Pseudo-F* dapat digunakan untuk menentukan berapa banyak cluster yang ideal yang akan terbentuk. [2]. *Pseudo-F* sebagai ukuran kualitas pengelompokan didefinisikan sebagai rasio varians antara klaster terhadap varians dalam klaster. Nilai *Pseudo-F* yang lebih tinggi berarti pusat-pusat kelompok (cluster) terpisah lebih baik dan varians dalam kelompok relatif lebih kecil, mengindikasikan pengelompokan yang lebih baik dan pemisahan yang lebih baik antara klaster. Nilai *Pseudo-F* dirumuskan sebagai berikut :

$$Pseudo F = \frac{(R^2/c - 1)}{(1 - R^2/n - c)} \quad (7)$$

R^2 R Squared

c Jumlah kelompok yang terbentuk selama pengamatan

n Jumlah objek yang diamati

Setelah jumlah *cluster* optimum didapatkan dalam setiap *link*, kemudian menentukan metode *link* terbaik berdasarkan nilai Internal Cluster Dispersion (*ICD*). *ICD Rate* merupakan sebuah metrik dalam analisis klaster untuk mengevaluasi seberapa padat atau tersebar data di dalam setiap kelompok (cluster) yang terbentuk [13]. *ICD Rate* memberikan informasi mengenai tingkat kepadatan atau kerapatan objek di dalam suatu kelompok. Semakin rendah nilai *ICD Rate*, semakin baik kelompok tersebut karena menunjukkan bahwa objek-objek di dalam klaster lebih mendekati pusat kelompok atau satu sama lain. Nilai *ICD Rate* dirumuskan sebagai berikut :

$$ICD Rate = 1 - \frac{SSB}{SST} = 1 - R^2 \quad (8)$$

2.8. Kinerja Hasil Pengelompokan Kategorik

kinerja hasil pengelompokan variabel kategorik dilihat dari ANOVA (*analysis of variance*). Semakin kecil nilai rasio antara simpangan baku di dalam *cluster* (S_W) dan simpangan baku antar *cluster* (S_B), semakin baik kinerja metode pengelompokan. Ini menunjukkan bahwa setiap cluster memiliki tingkat homogenitas maksimum, dan setiap cluster memiliki tingkat heterogenitas maksimum.[7]. Simpangan baku di dalam *cluster* (S_W) dan simpangan baku antar *cluster* (S_B) dirumuskan pada persamaan berikut :

$$S'_W = [MSW]^{\frac{1}{2}} \quad (9)$$

$$S'_B = [MSB]^{\frac{1}{2}} \quad (10)$$

MSW *Mean of Squares Within*

MSB *Mean of Squares Between*

2.9. Indeks Desa Membangun (IDM)

Dalam Buku Rencana Pembangunan Jangka Menengah Nasional (RPJMN) dijelaskan bahwa IDM dikembangkan agar memperkuat capaian target pembangunan Desa. IDM bertumpu terhadap tiga Indeks, diantaranya Ketahanan Sosial, Ketahanan Ekonomi, dan Ketahanan Lingkungan [17]. Permendesa Nomor 2 tahun 2016 mengklasifikasikan 5 peringkat status desa IDM yaitu Desa Mandiri, Desa Maju, Desa Berkembang, Desa Tertinggal, dan Desa sangat Tertinggal[12].

3. METODE PENELITIAN

Penelitian ini menggunakan data sekunder dari hasil pendataan 277 Desa Potensi Desa (PODES) Kabupaten Banjar tahun 2021 oleh lembaga Badan Pusat Statistik . Terdapat 18 variabel numerik dan 29 variabel kategorik yang digunakan sesuai indikator IDM. Uraian prosedur penelitian sebagai berikut [4] & [13]:

- 1) Gambaran Karakteristik Desa
 1. Analisis data deskriptif variabel numerik menggunakan nilai rata – rata, nilai maksimal dan nilai minimal.
 2. Analisis data deskriptif variabel kategorik menggunakan nilai *modus*.
 - 2) Pengelompokan Data Tipe Numerik
 1. Pengelompokan dilakukan dengan metode hirarki *agglomerative single linkage*, *complete linkage*, dan *average linkage* dan menggunakan jarak *Euclidean*.
 2. Menghitung nilai *Pseudo F* dan *ICD Rate* dalam penentuan *Link* terbaik.
 - 3) Pengelompokan Data Tipe Kategorik
 1. Melakukan pengelompokan menggunakan metode QROCK dengan nilai *threshold* (θ) antara 0 dan 1
 2. Menghitung rasio S_W dan S_B guna menentukan jumlah kluster optimum.
 - 4) *Final Cluster* Pengelompokan Metode Ensemble QROCK
 1. Menggabungkan hasil dari pengelompokan data numerik dan data kategorik.
 2. Mengelompokkan *cluster* gabungan menggunakan metode QROCK dengan nilai *threshold* (θ) antara 0 dan 1 (metode Ensemble QROCK).
 3. Menghitung rasio S_W dan S_B guna menentukan jumlah kluster ideal sebagai kluster akhir.
 - 5) Interpretasi, Kesimpulan, dan Saran.
-

4. HASIL DAN PEMBAHASAN

4.1 Pengelompokan Data Numerik

4.1.1. Kinerja Pengelompokan Data Numerik

Mendapatkan jumlah kelompok ideal dari tiga metode tautan yaitu *single linkage*, *complete linkage*, dan *average linkage* diperoleh dari Nilai *Pseudo F* maksimum untuk masing masing metode tautan.

Tabel 1. Nilai Pseudo-F

Linkage	Jumlah Cluster	Pseudo-F
single	2	185,6792
	3	94,585
	4	62,9547
	5	63,7107
	6	50,8406
	7	44,4873
	complete	2
3		293,6048
4		690,6479
5		567,2634
6		478,8821
7		599,0722
average	2	185,6792
	3	293,6048
	4	199,932
	5	449,8602
	6	382,269
	7	331,6747

Jumlah *cluster* optimum masing – masing metode linkage diperoleh berdasarkan nilai pseudo-f maksimum yaitu *single linkage* sebanyak 2 *cluster*, *complete linkage* sebanyak 4 *cluster*, dan *average linkage* sebanyak 5 *cluster*. Selanjutnya, membandingkan ketiga metode linkage berdasarkan nilai *ICD rate* minimum untuk mengidentifikasi metode optimal pengelompokan desa – desa di Kabupaten Banjar.

Tabel 2. Nilai ICD Rate

Linkage	Jumlah Cluster	ICD Rate
single	2	0,5969
complete	4	0,1164
average	5	0,1313

Berdasarkan tabel diatas, Metode linkage adalah complete linkage dengan membentuk 4 *cluster* dengan nilai ICD Rate minimum sebesar 0,1164.

4.2 Pengelompokan Data Kategorik

4.2.1. Kinerja Pengelompokan Data Kategorik

Data kategorik dikelompokkan menggunakan metode *QROCK* menggunakan nilai threshold sebagai batas penentuan tetangga, dengan nilai - nilai 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, dan 0.9. Nilai terkecil rasio S_W dan S_B dipilih dari sepuluh nilai ambang batas yang telah ditentukan.

Tabel 3. Rasio S_W dan S_B Pengelompokan kategorik

Threshold	Jumlah kelompok	S_W	S_B	Ratio
0,1	1	1,76E-09	Inf	0
0,2	1	1,76E-09	Inf	0
0,3	1	1,76E-09	Inf	0
0,4	1	1,76E-09	Inf	0
0,5	1	1,76E-09	Inf	0
0,6	3	8,26E+04	1,426	0,057
0,7	18	1,67E+06	9,041	0,184
0,8	116	2,53E+07	38,81	0,652
0,9	259	1,75E+08	62,02	2,824

nilai rasio antara S_W dengan S_B terkecil berada pada treshold 0,6 yakni 0,057. Pengelompokan desa di Kabupaten Banjar variabel kategorik menghasilkan jumlah *cluster* sebanyak 3 *cluster*.

4.3 Pengelompokan Data Campuran

Hasil pengelompokan data numerik dan data kategorik yang telah diperoleh kemudian dilakukan pengelompokan dengan memandang hasil sebagai data kategorik baru.

Tabel 4. Data Campuran hasil pengelompokan numerik dan kategorik

No	Desa	Hasil kelompok numerik	Hasil kelompok kategorik
1	Sungai Musang	1	1
2	Bakambat	1	1
3	Labat Muara	1	1
4	Tanipah	2	1
5	Pemurus	3	1
6	Bunipah	1	1
.	.	.	.
.	.	.	.
.	.	.	.
272	Karya Makmur	1	1
273	Makmur Karya	1	1
274	Alalak Padang	1	1
275	Benua Anyar	1	1

No	Desa	Hasil kelompok numerik	Hasil kelompok kategorik
276	Garis Hanyar	1	3
277	Sumber Sari	1	1

4.3.1. Kinerja Pengelompokan Data Ensemble QROCK

Data hasil gabungan dikelompokkan menggunakan metode *QROCK*. Inisiasi threshold sebagai ambang batas tetangga, dengan nilai-nilai yaitu 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, dan 0.9. Nilai terkecil rasio S_W dan S_B dipilih dari sepuluh nilai ambang batas yang telah ditentukan.

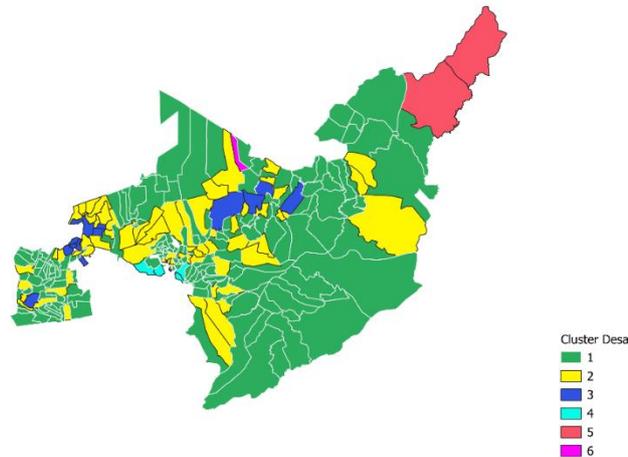
Tabel 5. Rasio S_W dan S_B Pengelompokan Ensemble QROCK

Treshold	Jumlah kelompok	S_W	S_B	Ratio
0,1	1	2,11E-09	Inf	0
0,2	1	2,11E-09	Inf	0
0,3	1	2,11E-09	Inf	0
0,4	6	4,82E-09	1	4,82E-09
0,5	6	4,82E-09	1	4,82E-09
0,6	6	4,82E-09	1	4,82E-09
0,7	6	4,82E-09	1	4,82E-09
0,8	6	4,82E-09	1	4,82E-09
0,9	6	4,82E-09	1	4,82E-09

Dari tabel 5 nilai rasio treshold 0,1 – 0,3 tidak dipilih sebagai ambang batas yang digunakan karena hanya menghasilkan 1 cluster dan nilai rasio adalah 0,000, meskipun lebih kecil dari treshold 0,4 - 0,9. pengelompokan desa di Kabupaten Banjar data hasil gabungan menghasilkan jumlah *cluster* sebanyak 6 *cluster*.

4.4 Karakteristik Desa Hasil Pengelompokan Ensemble QROCK

Dengan menggunakan metode ensemble QROCK, desa untuk Kabupaten Banjar terbagi menjadi enam kelompok. Adapun persebaran desa-desa menurut kelompok - kelompoknya dapat dilihat pada peta clusterisasi IDM Desa Kabupaten Banjar tahun 2021 ditunjukkan pada gambar 1.



Gambar 1. Peta clusterisasi Ensemble QROCK Desa Kabupaten Banjar tahun 2021

Karakteristik setiap kelompok desa memberikan gambaran lebih lanjut tentang perbedaan karakteristik kelompok desa Kabupaten Banjar tahun 2021. Untuk memudahkan pemahaman tentang karakteristik desa setiap kelompok, maka uraian deskripsi karakteristik desa – desa Kabupaten Banjar dibagi berdasarkan masing – masing dimensi dan sub dimensi Indeks Desa Membangun (IDM).

4.4.1 Karakteristik Desa Kelompok 4

Kelompok desa yang paling baik karakteristiknya di Kabupaten Banjar berada di kelompok 4 yang terdiri dari 4 desa. Dari segi pendidikan masing – masing desa memiliki akses ke pendidikan paud, dasar, dan sekolah menengah. Hal ini menandakan bahwa desa yang berada di kelompok 4 terpenuhi warganya untuk mendapatkan akses wajib belajar 12 tahun. Dari segi kesehatan desa kelompok 4 tersedia pelayanan kesehatan dasar bagi masyarakatnya diiringi dengan jumlah tenaga kesehatan, bidan, dokter, dan nakes yang sangat tersedia jumlahnya dengan rata – rata 20 jumlah tenaga. Dari segi akses pusat perdagangan dan pasar desa kelompok 4 sudah memiliki akses untuk pertokoan dan pasar, dan juga didukung oleh jumlah warung dan minimarket yang banyak pada masing – masing desa. akses air bersih sumber mandi dan mencuci mayoritas penduduk desa kelompok 4 menggunakan sumur bor atau pompa, dan sumber air minum mayoritas menggunakan air isi ulang

4.4.2 Karakteristik Desa Kelompok 3

Kelompok desa yang baik karakteristiknya di Kabupaten Banjar tergolong ke dalam kelompok 3 yang terdiri dari 14 desa. Segi pendidikan memiliki akses ke pendidikan dasar dan paud, dan rata – rata desa kelompok 3 memiliki akses ke pendidikan sekolah menengah pertama. Dari segi kesehatan desa kelompok 3 tersedia pelayanan kesehatan dasar. segi akses pusat perdagangan dan pasar desa kelompok 3 sudah memiliki akses untuk pertokoan dan pasar. akses air bersih sumber mandi dan mencuci mayoritas penduduk desa kelompok 3 menggunakan air ledeng meteran (PAM/PDAM), dan sumber air minum mayoritas menggunakan air kemasan bermerek.

4.4.3 Karakteristik Desa Kelompok 2

Kabupaten Banjar memiliki sebanyak 70 desa yang tergolong ke dalam kelompok 2 berdasarkan hasil pengelompokan. Segi pendidikan masing – masing desa memiliki akses ke pendidikan dasar dan paud. Dari segi kesehatan desa kelompok 2 masing – masing sudah tersedia pelayanan kesehatan dasar bagi masyarakat. Segi akses pusat perdagangan dan pasar desa kelompok 2 masih minim aksesnya untuk pertokoan dan pasar. Akses air bersih penduduk desa kelompok 2 beragam sumber mandi dan mencucinya, namun mayoritas menggunakan air sungai/ danau/ kolam/ waduk/ bendungan, Begitu pun sumber air minum mayoritas menggunakan air sungai/ danau/ kolam/ waduk/ bendungan.

4.4.4 Karakteristik Desa Kelompok 1

Kabupaten Banjar memiliki sebanyak 186 desa yang tergolong ke dalam kelompok 1 berdasarkan hasil pengelompokan. Segi pendidikan masing – masing desa memiliki akses ke pendidikan dasar dan paud. Dari segi kesehatan desa kelompok 1 masing – masing sudah tersedia pelayanan kesehatan dasar bagi masyarakat. Dari segi akses pusat perdagangan dan pasar desa kelompok 1 masih minim aksesnya untuk pertokoan dan pasar. akses air bersih penduduk desa kelompok 1 beragam sumber mandi dan mencucinya, namun mayoritas menggunakan air sungai/ danau/ kolam/ waduk/bendungan. Sedangkan sumber air minum masih menggunakan air sumur.

4.4.5 Karakteristik Desa Kelompok 5

Kabupaten Banjar memiliki sebanyak 2 desa yang tergolong ke dalam kelompok 5 berdasarkan hasil pengelompokan. segi pendidikan masing – masing desa memiliki akses ke pendidikan dasar dan paud. Dari segi kesehatan desa kelompok 5 minim pelayanan kesehatan dasar bagi masyarakatnya. Dari segi akses pusat perdagangan dan pasar desa kelompok 5 masih tidak memiliki akses. akses air bersih penduduk desa kelompok 5 sumber mandi dan mencucinya menggunakan mata air dan sungai/kolam/waduk . Sedangkan sumber air minum menggunakan mata air.

4.4.6 Karakteristik Desa Kelompok 6

Kelompok enam merupakan *outlier* dari hasil keseluruhan pengelompokan dikarenakan hanya memiliki sebanyak 1 desa yang tergolong ke dalam kelompok 6.

5. KESIMPULAN

Pengelompokan desa di Kabupaten Banjar menggunakan metode Ensemble QROCK menghasilkan 6 kelompok. Kelompok 4 menonjol dengan fasilitas lengkap dan lingkungan baik. Kelompok 3 memiliki karakteristik serupa namun perlu peningkatan ekonomi. Kelompok 2 memiliki akses dasar dengan perhatian khusus pada infrastruktur logistik. Kelompok 1 memiliki akses dasar namun kekurangan pada pusat perdagangan dan perbankan.

Desa yang perlu diprioritaskan adalah Kelompok 5, dan Kelompok 6 sebagai outlier. Rekomendasi pembangunan dapat disesuaikan dengan karakteristik masing-masing kelompok, memberikan dasar yang kuat bagi pemerintah desa dalam menetapkan kebijakan pembangunan yang efektif.

DAFTAR PUSTAKA

- [1] BPS Kabupaten Banjar (2022), *Statistik Daerah Kabupaten Banjar Tahun 2022*.
- [2] Caliński, T., and Harabasz, J. (1974), "A Dendrite Method For Cluster Analysis," *Communications in Statistics*, 3, 1–27. <https://doi.org/10.1080/03610927408827101>.
- [3] Danapriatna, N., and Setiawan, R. (2005), *Pengantar Statistika*, Yogyakarta: Graha Ilmu.
- [4] Dutta, M., Mahanta, A. K., and Pujari, A. K. (2005), "QROCK: A quick version of the ROCK algorithm for clustering of categorical data," *Pattern Recognition Letters*, 26, 2364–2373. <https://doi.org/10.1016/j.patrec.2005.04.008>.
- [5] Guha, S., Rastogi, R., and Shim, K. (1999), "ROCK: Robust Clustering Algorithm for Categorical Attributes," 512–521.
- [6] Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2009), *Multivariate Data Analysis*, New Jersey.
- [7] Hanandya, S., Nur A, M., Nur, I. M., and Haris, A. (2021), *Perbandingan Metode Single Linkage dan Average Linkage Pada Comparison Of Single Linkage and Average Linkage In The Grouping Of Sub-Districts In Kendal Based On The Potential Of Agriculture Sector*, Semarang.
- [8] He, Z., Xu, X. i, and Deng, S. (2002), *Clustering Mixed Numerical and Categorical Data : A Cluster Ensemble Approach*, Harbin.
- [9] Johnson, R. A., and Wichern, D. W. (2002), *Applied Multivariate Statistical Analysis*, New Jersey: Pearson Education, Inc.
- [10] KEMENTERIAN DESA, P. D. T. D. T. (2015), *INDEKS DESA MEMBANGUN 2015*, Jakarta.
- [11] Lubis, I. A., Nugroho, S., and Swita, D. B. (n.d.). "KAJIAN METODE PENGKLASTERAN HIRARKI DENGAN BERBAGAI PENGUKURAN JARAK," *e-Jurnal Statistika*, 42–53.
- [12] MENTERI DESA, P. D. T. D. T. R. I. (2016), *Permendesa Nomor 2 Tahun 2016*.
- [13] Ramdhany, R. (2018), "PENGELOMPOKAN DESA DI KABUPATEN BONDOWOSO BERDASARKAN DATA CAMPURAN NUMERIK DAN KATEGORIK MENGGUNAKAN METODE ENSEMBEL ROCK," Surabaya: INSTITUT TEKNOLOGI SEPULUH NOPEMBER.
- [14] Ronald E. Walpole (1995), *Pengantar statistika*, Jakarta: PT Gramedia Pustaka Utama.
- [15] Sharma, S. (1996), *Applied Multivariate Techniques*, New York: John Wiley and Sons.
- [16] Sharpe, N. R., De Veaux, R. D., and Velleman, P. F. (2018), *Business statistics*.

-
- [17] Sihombing, P. R., Arsani, A. M., Angraini, W., and Wisnu Pratiko (2022), "IMPLEMENTASI MODEL SEM PADA HUBUNGAN IPM, IPD DAN IDM," *Jurnal Bayesian : Jurnal Ilmiah Statistika dan Ekonometrika*, 2, 107–115. <https://doi.org/10.46306/bay.v2i2>.
- [18] Späth, H. (1982), *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*, New York: Halsted Press.
- [19] Suguna, J., and Arul Selvi, M. (2012), "Ensemble Fuzzy Clustering for Mixed Numeric and Categorical Data," *International Journal of Computer Applications*, Foundation of Computer Science, 42, 19–23. <https://doi.org/10.5120/5672-7705>.