



Pengelompokan Provinsi di Indonesia Berdasarkan Produksi dan Nelayan Cakalang 2023 dengan *Block-Based K-Medoids*

Khansa Qamila¹, Arum Handini Primandari²,^{1,2} Program Studi Statistika, Universitas Islam Indonesia² primandari.arum@uii.ac.idCorresponding author email: 22611107@students.uui.ac.id

Abstract: The existence of thousands of islands makes Indonesia a country with abundant marine resources and great strategic potential. Skipjack tuna is one of the leading commodities that dominates the capture fisheries sector in Indonesia. However, its utilization is still uneven across provinces, as reflected in disparities in production volume, production value, and the number of fishermen. This study focuses on classifying provinces in Indonesia based on similarities across these three indicators in 2023 using the Block-Based K-Medoids approach, a method that selects representative objects based on block divisions formed from standard deviation and total variable values to determine initial medoids, thereby ensuring greater stability and accuracy in generating clusters. The analysis results show 3 provincial clusters with different characteristics, namely cluster 1 has low production volume and value but high number of fishermen, indicating low productivity. Cluster 2 shows moderate production volume and value with the least number of fishermen, reflecting high productivity efficiency. Meanwhile, cluster 3 has the highest volume, production value, and number of fishers, reflecting a region with a large and developed capture fisheries sector.

Keywords: Capture Fisheries, Skipjack, Cluster, Block-Based K-Medoids, Fisheries Production.

Abstrak: Keberadaan ribuan pulau menjadikan Indonesia sebagai negara dengan kekayaan sumber daya kelautan yang melimpah dan potensi strategis yang besar, Cakalang termasuk dalam deretan komoditas unggulan yang mendominasi sektor perikanan tangkap di Indonesia. Namun, pemanfaatannya masih belum merata di seluruh provinsi, yang tercermin dari ketimpangan volume produksi, nilai produksi, dan jumlah nelayan. Studi ini difokuskan pada klasifikasi provinsi di Indonesia dengan mempertimbangkan kesamaan berdasarkan ketiga indikator tersebut pada tahun 2023 menggunakan pendekatan Block-Based K-Medoids, yaitu metode yang memilih objek representatif berdasarkan pembagian blok yang dibentuk dari standar deviasi dan total nilai variabel. untuk menentukan medoid awal, sehingga lebih stabil dan akurat dalam menghasilkan kelompok. Hasil analisis menunjukkan 3 *cluster* provinsi dengan karakteristik berbeda yaitu *cluster* 1 memiliki volume dan nilai produksi rendah namun jumlah nelayan tinggi, mengindikasikan produktivitas rendah. *Cluster* 2 menunjukkan volume dan nilai produksi sedang dengan jumlah nelayan paling sedikit, mencerminkan efisiensi produktivitas yang tinggi. Sedangkan, *cluster* 3 memiliki volume, nilai produksi, dan jumlah nelayan tertinggi, mencerminkan daerah dengan sektor perikanan tangkap yang besar dan berkembang.

Kata kunci: Perikanan Tangkap, Ikan Cakalang, Block-Based K-Medoids, Produksi Perikanan.

I. PENDAHULUAN

Letak Indonesia sebagai negara kepulauan yang dikelilingi perairan luas menjadikannya kaya akan potensi sumber daya alam kelautan. Salah satu komoditas utamanya yaitu cakalang, menurut [1]. Ikan Cakalang adalah salah satu spesies ikan yang memiliki peran signifikan dalam perekonomian serta memiliki berbagai manfaat bagi nelayan. Hal ini sejalan dengan temuan dari [2] yang menyebutkan bahwa cakalang memiliki nilai yang cukup tinggi dari produksi maupun ekonominya. Hal tersebut dikarenakan penangkapan ikan cakalang secara umum dapat dilakukan sepanjang tahun dan harganya yang termasuk stabil di pasaran. Maka dari itu cakalang memiliki kontribusi yang besar dalam peningkatan kesejahteraan masyarakat Indonesia khususnya bagi mereka yang bermata pencaharian sebagai nelayan.

Meskipun ikan cakalang memiliki potensi besar dalam sektor perikanan, pemanfaatannya masih belum merata di seluruh provinsi di Indonesia. Data KKP tahun 2023 menunjukkan adanya perbedaan signifikan antar daerah dalam hal volume produksi, nilai produksi, dan jumlah nelayan yang terlibat. Menurut [2], terdapat provinsi dengan produksi tinggi namun nilai ekonominya belum optimal. Kondisi ini menunjukkan bahwa pengelolaan perikanan cakalang belum sepenuhnya efisien, sehingga



diperlukan analisis lebih lanjut untuk memahami karakteristik tiap wilayah dan mendorong pengelolaan yang lebih merata.

Berdasarkan permasalahan ketimpangan dalam pemanfaatan potensi ikan cakalang di Indonesia, tujuan dari studi ini adalah untuk melakukan klasifikasi terhadap provinsi-provinsi dengan mengacu pada tiga indikator utama: volume produksi, nilai produksi, dan jumlah nelayan ikan cakalang pada tahun 2023 menggunakan metode *K-Medoids*. Pemilihan metode ini didasarkan pada kemampuannya dalam mengelola data yang mengandung *outlier* dan menghasilkan *cluster* yang lebih stabil dan representatif. Pengelompokan ini diharapkan dapat membantu memahami lebih jauh mengenai pola distribusi sektor perikanan cakalang di Indonesia, serta memberikan dasar yang kuat bagi pemerintah untuk merumuskan kebijakan pembangunan perikanan yang lebih efisien, merata, dan berkelanjutan.

Penelitian sebelumnya oleh [3] dalam *Clustering* Produksi Perikanan Budidaya Laut Berdasarkan Provinsi Menggunakan Algoritma *K-Means* telah berhasil mengelompokkan provinsi di Indonesia berdasarkan produksi perikanan budidaya laut menggunakan algoritma *K-Means*. Namun, penelitian tersebut lebih fokus pada perikanan budidaya, sedangkan penelitian ini akan mengaplikasikan metode *Block Based K-Medoids* untuk mengelompokkan provinsi berdasarkan data perikanan tangkap, khususnya ikan cakalang. Perbedaan utama terletak pada objek yang dianalisis dan metode yang digunakan, di mana *K-Medoids* lebih tahan terhadap *outlier* dan lebih stabil dalam menangani data yang memiliki variasi ekstrem, seperti yang sering terjadi pada sektor perikanan tangkap. Oleh karena itu, hasil yang diperoleh diharapkan dapat memberikan wawasan yang lebih komprehensif dalam pengelompokan sektor perikanan cakalang di Indonesia dan memberikan dasar yang lebih kuat untuk pengambilan keputusan yang lebih tepat sasaran dalam kebijakan perikanan

II. METODE PENELITIAN

2.1. Data dan Sumber Data

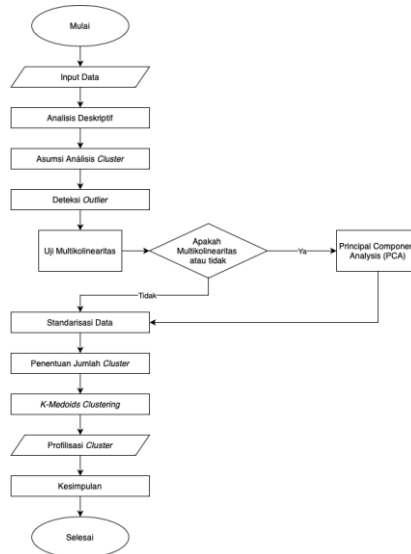
Sumber data dalam penelitian ini berasal dari data sekunder yang diakses melalui laman resmi <https://portaldata.kkp.go.id/portals/data-statistik/nelayan/tbl-statis/d/65> data tersebut memuat informasi mengenai volume produksi, nilai produksi, dan jumlah nelayan ikan cakalang tahun 2023 yang tercatat pada masing-masing provinsi di Indonesia. Pada penelitian ini menggunakan 4 variabel dengan rincian seperti pada tabel berikut.

Tabel 1. Variabel

Nama Variabel	Definisi Operasional
Provinsi	Wilayah administrasi tempat aktivitas penangkapan ikan cakalang dilaporkan.
Volume produksi	Jumlah hasil tangkapan ikan cakalang (ton).
Nilai produksi	Nilai ekonomi hasil tangkapan ikan cakalang (Rp).
Jumlah nelayan	Total nelayan yang terlibat dalam penangkapan ikan cakalang.

2.2. Alur Penelitian

Studi ini mengombinasikan analisis deskriptif dan metode *Block Based K-Medoids* dalam pengolahan data, dengan analisis deskriptif berfungsi memberikan pemahaman awal terhadap data. Sementara itu, metode *Block Based K-Medoids Clustering* digunakan untuk melakukan pengelompokan terhadap objek-objek dalam suatu *dataset* yang berpotensi mengandung *outlier* [4]. Penggunaan metode ini bertujuan untuk mengidentifikasi pola distribusi sektor perikanan cakalang secara lebih akurat.



Gambar 1. Diagram Alir Penelitian

2.3. Statistika Deskriptif

Statistika Deskriptif bertujuan untuk merangkum dan memvisualisasikan data yang dimiliki agar lebih mudah dipahami dan di analisis [5]. Data yang dianalisis menggunakan statistik deskriptif umumnya ditampilkan melalui tabel, grafik, diagram, maupun pictogram, disertai dengan perhitungan ukuran pemusatan seperti mean, median, modus, kuartil, desil, dan presentil. Sementara itu, untuk menggambarkan distribusi data, digunakan ukuran penyebaran seperti standar deviasi, varians, dan persentase. Teknik analisis ini bertujuan untuk menggambarkan karakteristik data secara ringkas dan informatif.

2.4. Uji Multikolinearitas

Multikolinearitas adalah kondisi yang muncul ketika terdapat hubungan linear yang sangat erat atau sempurna antara sebagian atau seluruh variabel dalam suatu model [6]. Penelitian ini menerapkan *Variance Inflation Factor* (VIF) untuk mendeteksi multikolinearitas, dengan tujuan untuk mengukur tingkat peningkatan varians koefisien akibat interkorelasi antar variabel independen. Adapun rumus VIF adalah:

$$VIF = \frac{1}{1 - R^2} \quad (1)$$

Dalam hal ini, R^2 merupakan koefisien determinasi dari regresi variabel independen terhadap variabel independen lainnya [7]. VIF dengan nilai di atas 10, mengindikasikan adanya multikolinearitas yang cukup kuat antar variabel independen, yang berpotensi mengganggu keakuratan hasil analisis regresi.

2.5. Euclidean Distance

Menurut [8] *euclidean distance* bertujuan untuk mengetahui jarak linear antara dua titik dalam ruang *euclidean*, dengan mempertimbangkan hubungan antara jarak dan sudut. Pada bidang matematika konsep ini serupa dengan rumus pythagoras menghitung jarak antara dua titik dalam satu dimensi [9].



$$d_{ij} = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (2)$$

Adapun nilai $i = 1, 2, \dots, n; j = 1, 2, \dots, n; l = 1, 2, \dots, p$ dan d_{ij} adalah jarak objek i dan objek j .

2.6. Penentuan Jumlah Cluster Optimal

Penentuan jumlah k cluster optimal dilakukan dengan bantuan metode *elbow* dan *silhouette*. Metode *Elbow* digunakan dengan cara membandingkan persentase antar cluster. Hasil analisis dengan metode ini divisualisasikan dalam bentuk grafik [10]. Perbandingan tersebut dihitung dengan menggunakan nilai *Sum of Square Error* (SSE) sebagai berikut.

$$SSE = \sum_{k=1}^k \sum_{x_i \in C_k} |x_i - C_k|^2 \quad (3)$$

Dalam hal ini k mewakili jumlah cluster, x_i mengacu pada data pada urutan ke- i , C_k merujuk pada Cluster ke- k , dan C_k adalah pusat atau *centroid* dari cluster ke- k .

Silhouette merupakan metrik yang mengukur tingkat kesamaan data dalam suatu cluster, dengan penghitungan yang dilakukan secara individual untuk setiap objek dalam kelompok tersebut. Rumus *silhouette* sebagai berikut.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

Nilai *Silhouette* yang hamper mendekati 1 menunjukkan kualitas pengelompokan yang sangat baik pada kelompok tersebut. Sebaliknya, jika nilainya mendekati -1, kualitas pengelompokan dianggap buruk pada Cluster tersebut [11].

2.7. Block-Based K-Medoids Clustering

Pada metode *Block-Based K-Medoids*, pemilihan medoid awal dilakukan dengan menggunakan objek representatif yang berasal dari blok berdasarkan standar deviasi serta jumlah nilai dari masing-masing variabel. Untuk mengimplementasikan metode ini, beberapa parameter utama perlu ditetapkan sebelumnya guna memastikan akurasi dan efektivitas pengelompokan [12]. Jika terdapat n objek dengan p variabel yang berupa numerik, kategorik, atau campuran, maka perhitungan standar deviasi untuk objek ke- i mencakup semua nilai pada p variabel dapat dihitung menggunakan formula berikut:

$$\mu_i = \frac{\sum_{l=1}^p \frac{(x_{il} - \bar{x}_i)^2}{p}}{(p-1)} \quad (5)$$

Dalam hal ini, $\bar{x}_i = \frac{\sum w_i}{p}$, dimana w_i menunjukkan jumlah nilai dari seluruh p variabel atau dapat dijelaskan sebagai berikut:

$$w_i = \sum_{l=1}^p x_{il} \quad (6)$$

Dalam hal ini, i dan l masing-masing merepresentasikan indeks objek, yaitu dari 1 hingga nm . Parameter ini berperan sebagai panduan dalam menentukan medoid awal.



Untuk objek ke- l , rata-rata jarak terhadap *cluster* ke- g , yang dimiliki n_g , dinyatakan sebagai berikut:

$$b_i = \frac{1}{n_g} \sum_{j=1}^{n_g} d_{ij} \quad (7)$$

Jumlah jarak dari objek ke *medoid* dalam kluster ke- k disebut sebagai $TD(k)$ dinyatakan sebagai berikut:

$$TD(k) = \sum_{g=1}^k \sum_{i=1}^{n_g} \sum_{l=1}^p |x_{gil} - m_{gl}| \quad (8)$$

di mana, x_{gil} merupakan nilai objek ke- i untuk variabel ke- l pada *cluster* ke- g dan m_{gl} adalah medoid dari kluster ke- g untuk variabel ke- l .

Berikut merupakan tahapan dalam penerapan algoritma Block-Based K-Medoids [13] :

Tahap 1: Pemilihan objek representatif awal

- Pada tahap ini, untuk setiap objek i , dengan $i = 1, 2, \dots, n$, dihitung dua parameter utama, yaitu standar deviasi (μ_i) dan jumlah nilai variabel (w_i).
- Selanjutnya, seluruh objek diurutkan berdasarkan nilai (μ_i) secara menaik. Jika terdapat beberapa objek dengan nilai standar deviasi yang sama, maka pengurutan dilanjutkan berdasarkan nilai (w_i) secara menaik.
- Dari hasil pengurutan tersebut, dipilih k blok awal berdasarkan nilai μ_i dan w_i (atau hanya berdasarkan μ_i jika tidak dikombinasikan), objek pertama dari masing-masing blok tersebut kemudian ditetapkan sebagai *medoid* awal.
- Langkah terakhir pada tahap ini adalah menetapkan keanggotaan awal setiap objek ke dalam salah satu dari k kluster berdasarkan jarak terdekat antara objek dan medoid.

Tahap 2: pembentukan partisi himpunan data

- Pada tahap ini, medoid pada masing-masing kluster diperbarui dengan memilih objek yang menghasilkan nilai rata-rata jarak terkecil terhadap seluruh anggota dalam kluster tersebut.
- Setelah medoid diperbarui, dilakukan identifikasi ulang keanggotaan kluster dengan menugaskan setiap objek ke medoid terdekat. Kemudian, dihitung total jarak simpangan semua objek terhadap medoidnya, yang dinotasikan sebagai $TD(k)$

Proses ini diulang secara iterative hingga nilai $TD(k)$ tidak mengalami perubahan dibandingkan dengan iterasi sebelumnya, jumlah iterasi maksimum telah dicapai, atau komposisi medoid tidak berubah.

III. HASIL DAN PEMBAHASAN

3.1. Analisis Deskriptif

Analisis Deskriptif digunakan dalam penelitian ini guna memberikan pemahaman umum terhadap data, melalui pengamatan terhadap nilai maksimal (*max*), nilai-nilai minimum (*min*), nilai rata-rata (*mean*) dan nilai median [14] . Disajikan pada Tabel 2 berikut ini.

Tabel 2. Statistika Deskriptif

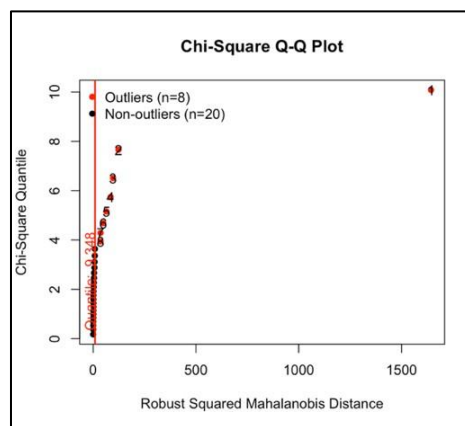
Variabel	Minimum	Maksimum	Mean	Median
Volume Produksi	0.32	71349.81	16879.72	8368.11
Nilai Produksi	2916	1391249873	3503	1758
Jumlah Nelayan	5253	242464	87854	73741



Volume Produksi berkisar antara 0.32 hingga 71349.81 ton dengan rata-rata 16879.72 ton. Nilai Produksi (Rp) berkisar dari Rp 2.916.000 hingga Rp 1.391.000.000 dengan rata-rata Rp 350.300.000. Jumlah Nelayan tercatat antara 5253 hingga 242464 orang dengan rata-rata 87854 orang. Ketiga variabel menunjukkan sebaran data yang lebar dan kecenderungan distribusi condong ke kanan. Dalam *Block Based K-Medoids Clustering*, penentuan *medoid* menggunakan konsep median sebagai acuan, proses ini memungkinkan pemilihan titik data aktual yang paling mewakili setiap *cluster*, untuk mengurangi dampak dari *outlier* atau nilai ekstrem. Hal ini penting untuk memastikan bahwa *cluster* yang terbentuk dapat mencerminkan distribusi data secara lebih akurat.

3.2. Pengecekan Outlier

Deteksi *outlier* dilakukan untuk mengidentifikasi data yang menyimpang dari sebagian besar data lainnya. Meskipun *outlier* sering dianggap sebagai gangguan karena memiliki pola yang berbeda, data tersebut sebenarnya bisa mengandung informasi yang sangat berharga. Oleh karena itu, peneliti melakukan pengecekan *outlier* secara multivariat menggunakan metode kuantil dan didapatkan visualisasi sebagai berikut.



Gambar 2. Deteksi *Outlier*

Berdasarkan Gambar 2. data yang ditandai dengan delapan titik merah merupakan *outlier*, sedangkan titik-titik hitam merepresentasikan data yang bukan *outlier*. Dengan kondisi tersebut, metode *clustering* yang dipilih sebagai pendekatan paling sesuai dalam penelitian ini adalah *Block-Based K-Medoids*. Dalam metode ini, pusat *cluster* diwakili oleh objek representatif yang disebut *medoid*. *Medoid* tersebut kemudian menjadi acuan untuk mengelompokkan setiap objek ke dalam *cluster* terdekat. Karena karakteristik tersebut, *Block Based K-Medoids Clustering* sangat sesuai untuk diterapkan pada data yang mengandung *outlier*.

3.3. Uji Multikolinearitas

Multikolinearitas dapat diuji melalui analisis *Variance Inflation Factor* (VIF). nilai VIF yang berada dibawah atau sama dengan 10, maka dapat disimpulkan bahwa tidak terjadi multikolinieritas atau dengan kata lain memenuhi asumsi [15], berikut ini adalah nilai VIF dari variabel-variabel yang diteliti:

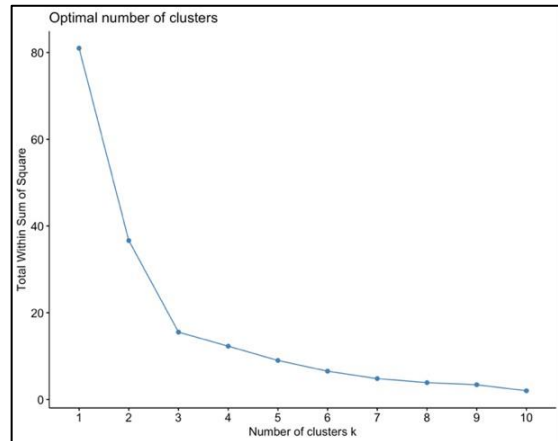
Tabel 3. *Output Uji Multikolinearitas*

Variabel	VIF
<i>XI</i>	7.193211
<i>XI</i>	7.221591
<i>XI</i>	1.079794

Hasil analisis menunjukkan bahwa semua variabel memiliki nilai VIF dibawah 10, mengindikasikan bahwa tidak terdapat korelasi atau multikolinearitas di antara variabel-variabel tersebut. Hal ini memastikan bahwa variabel-variabel tersebut independen dan memenuhi asumsi multikolinearitas.

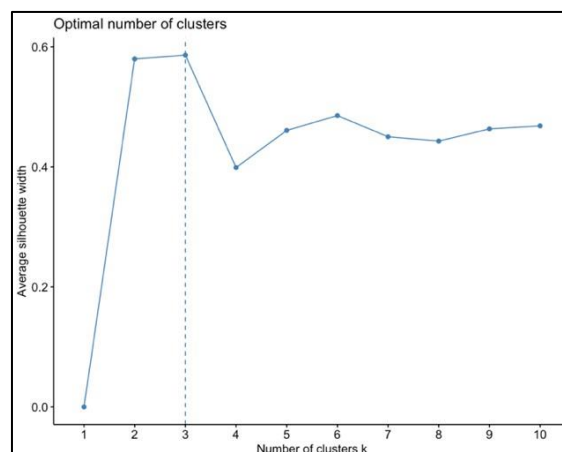
3.4. Penentuan Jumlah Cluster

Dalam menentukan jumlah *cluster* terbaik mencakup *Sum of Square Error* yang dikenal dengan pendekatan *Elbow* serta metode *Silhouette*.



Gambar 3. Grafik Optimal Metode *Elbow*

Dapat diamati dari Gambar 3. bahwa grafik mulai melandai pada saat $k=3$. Menurut [16] titik optimal jumlah klaster dapat ditentukan dari sudut siku yang terbentuk pada grafik. Dibandingkan nilai sebelumnya yang masih curam ($k = 1$ dan $k = 2$), kelandaian grafik ini menunjukkan bahwa tiga klaster merupakan pilihan optimal. Analisis selanjutnya dilakukan melalui metode *Silhouette*, dan hasilnya ditampilkan berikut ini.



Gambar 4. Grafik Optimal Metode *Silhouette*

Berdasarkan pengamatan pada Gambar 4. Metode *Silhouette* juga mengindikasikan bahwa nilai $k=3$ adalah jumlah *cluster* yang optimal. Konsistensi hasil antara metode *Elbow* dan *Silhouette* memperkuat dasar pemilihan tiga klaster dalam analisis ini, penulis menetapkan bahwa jumlah *cluster* optimal yang digunakan adalah 3 *cluster*.

3.5. Block-Based K-Medoids Clustering

Analisis *K-Medoids Clustering* dalam penelitian ini didukung oleh penggunaan *software* RStudio dan Microsoft Excel. RStudio digunakan untuk transformasi data, perhitungan jarak antar objek



menggunakan jarak *Euclidean*, serta pembuatan matriks jarak. Objek representatif awal (*medoid*) ditentukan berdasarkan kombinasi standar deviasi dan jumlah nilai yang diurutkan menaik. Setiap objek kemudian dikelompokkan ke dalam tiga *cluster* berdasarkan jarak terdekat ke *medoid*. Proses iterasi berlanjut dengan pemilihan *medoid* baru hingga total jarak simpangan tidak lagi berubah. Hasil dari pengelompokan menjadi 3 *cluster* disajikan dalam Tabel 4. di bawah ini.

Tabel 4. Hasil Medoid Akhir

Provinsi	$sd(u_i)$	$sum(w_i)$	Inisial	Iterasi			
			Grup	I-1	I-2	I-3	I-4**
Sulawesi Tenggara	15.1820	47.2637	G1*	G3	G3	G3	G3
Maluku Utara	17.3976	47.0532	G2*	G2	G2	G2	G2
Sulawesi Selatan	17.6422	62.7733	G3*	G3*	G3*	G3*	G3*
Papua	18.2201	50.2561	G2	G2	G2	G2	G2
Sumatera Utara	19.0058	56.5641	G3	G3	G3	G3	G3
Sumatera Barat	19.3158	42.3563	G2	G2	G2	G2	G2
Maluku	23.6523	73.4240	G3	G3	G3	G3	G3
Nusa Tenggara Barat	23.7471	41.5617	G1	G1	G1	G1	G1
Nusa Tenggara Timur	28.8301	42.4879	G1	G1	G1	G1	G1
DKI Jakarta	29.6567	88.6908	G2	G3	G3	G3	G3
Jawa Tengah	31.3975	76.2579	G3	G3	G3	G3	G3
Jawa Timur	32.3271	80.9465	G3	G3	G3	G3	G3
Bali	37.4968	40.6254	G2	G1	G1	G1	G1
Aceh	38.7321	82.9888	G2	G3	G3	G3	G3
Sulawesi Utara	38.7884	82.8418	G2	G3	G3	G3	G3
Jawa Barat	43.5611	44.0456	G1	G1*	G1	G1	G1
Gorontalo	45.7706	43.4684	G2	G2*	G2	G2	G2
Sulawesi Barat	46.0260	44.8603	G2	G2	G2*	G2*	G2*
Papua Barat	46.2770	46.0301	G2	G2	G2	G2	G2
Kalimantan Selatan	48.9494	43.8365	G1	G1	G1*	G1	G1
Kalimantan Timur	48.9562	43.3898	G1	G1	G1	G1*	G1*
Kalimantan Barat	49.8031	45.4351	G1	G1	G1	G1	G1
Sulawesi Tengah	52.0736	102.949	G3	G3	G3	G3	G3
Banten	53.7864	44.8433	G2	G2	G1	G1	G1
Lampung	58.2057	48.2709	G2	G2	G1	G1	G1
Bengkulu	58.2443	50.3146	G2	G2	G2	G2	G2
Kalimantan Utara	59.0196	49.5295	G2	G2	G2	G2	G2



Provinsi	$sd(u_i)$	$sum(w_i)$	Initial	Iterasi			
			Grup	I-1	I-2	I-3	I-4**
Daerah Istimewa							
Yogyakarta	59.1700	55.4490	G2	G2	G2	G2	G2
Total Jarak Simpangan Cluster			36.5604	26.0042	25.7432	25.7269	25.7269

*Medoid **Medoid Akhir

Pada iterasi ke-4, diperoleh hasil bahwa anggota *cluster* dan jumlah jarak simpangan tetap sama seperti pada iterasi ketiga, sehingga proses perhitungan dihentikan. *Medoid* akhir untuk $k=3$ ditetapkan pada Provinsi Kalimantan Timur, Sulawesi Barat, dan Sulawesi Selatan. Berdasarkan jarak masing-masing objek ke *medoid* akhir terdekat, peneliti mendapatkan anggota setiap *cluster* yaitu 9 provinsi dalam *cluster* 1, 9 provinsi dalam *cluster* 2 dan 10 provinsi dalam *cluster* 3. Nama-nama provinsi dari masing-masing anggota *cluster* disajikan pada Tabel 5. di bawah ini.

Tabel 5. Anggota Hasil *Cluster*

Cluster	Jumlah Cluster	Anggota Cluster
I	9	1) Bali
		2) Nusa Tenggara Timur
		3) Banten
		4) Kalimantan Timur
		5) Kalimantan Barat
		6) Jawa Barat
		7) Kalimantan Selatan
		8) Lampung
		9) Nusa Tenggara Barat
II	9	1) Bengkulu
		2) Daerah Istimewa Yogyakarta
		3) Kalimantan Utara
		4) Maluku Utara
		5) Papua
		6) Papua Barat
		7) Sulawesi Barat
		8) Sumatra Barat
		9) Gorontalo
III	10	1) Jawa Timur
		2) Maluku
		3) DKI Jakarta
		4) Sulawesi Selatan
		5) Sulawesi Tenggara
		6) Sulawesi Utara
		7) Sumatra Utara
		8) Jawa Tengah
		9) Sulawesi Tengah
		10) Aceh

3.6. Profilisasi *Cluster*

Setelah menentukan anggota tiap *cluster*, langkah selanjutnya adalah melakukan profilisasi data untuk menganalisis karakteristik setiap *cluster* melalui perhitungan median masing-masing variabel. Hasil profilisasi tersebut disajikan pada Tabel 5 berikut.



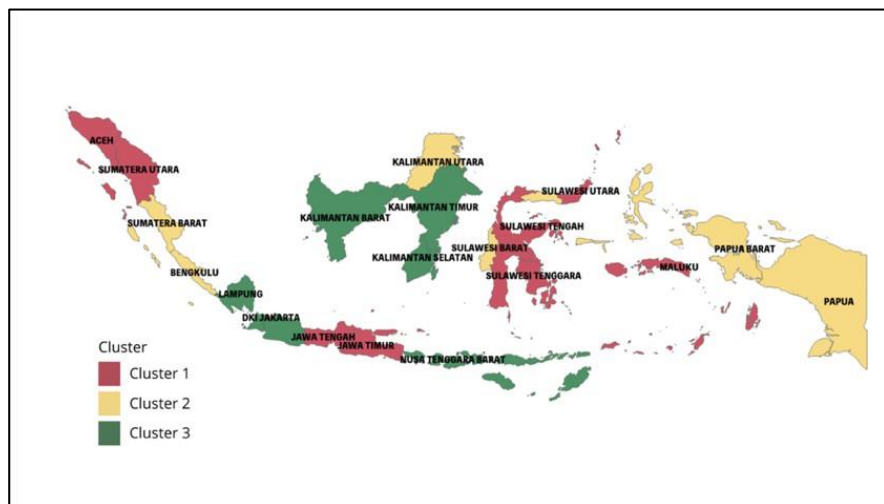
Tabel 6. Hasil Profilisasi *Clustering*

<i>Cluster</i>	Volume Produksi (ton)	Nilai Produksi (Rp)	Jumlah Nelayan
1	1979.50	37673502	71714
2	7877.62	18231965	28638
3	30051.91	655832546	140505

Berdasarkan nilai median pada setiap *cluster* yang terdapat pada Tabel 5 dapat diketahui karakteristik dari setiap *cluster* berdasarkan mediannya dapat dideskripsikan sebagai berikut:

- *Cluster 1* terdiri dari 9 provinsi yang ditandai dengan volume produksi (1979.50) dan nilai produksi (37673502) terendah, namun memiliki jumlah nelayan (71714) yang relatif tinggi. Hal ini mengindikasikan bahwa produktivitas per nelayan di *cluster* ini masih rendah.
- *Cluster 2* terdiri dari 9 provinsi, menunjukkan volume (7877.62) dan nilai produksi (18231965) sedang, namun dengan jumlah nelayan (28638) terendah di antara semua *cluster*. Karakter ini mencerminkan tingkat efisiensi atau produktivitas yang relatif tinggi, karena nilai dan volume produksi yang cukup besar dihasilkan oleh sedikit nelayan.
- *Cluster 3* mencakup 10 provinsi yang menunjukkan volume (30051.91) dan nilai produksi (655832546) tertinggi, sekaligus jumlah nelayan (140505) terbanyak. *Cluster* ini mencerminkan provinsi-provinsi dengan kapasitas produksi besar dan jumlah tenaga kerja tinggi, yang menunjukkan sektor perikanan tangkap yang sudah berkembang dan padat karya.

Hasil pengelompokan provinsi di Indonesia berdasarkan volume produksi, nilai produksi, dan jumlah nelayan ikan cakalang juga disajikan dalam bentuk peta seperti yang ditunjukkan pada Gambar 5. di bawah ini.



Gambar 5. Peta Persebaran *Cluster*

IV. KESIMPULAN

Berdasarkan rangkaian analisis serta uraian pembahasan sebelumnya, penulis dapat menyimpulkan bahwa dengan menggunakan metode *K-medoids clustering* penelitian ini berhasil mengelompokkan 28 provinsi di Indonesia ke dalam tiga *cluster* berdasarkan volume produksi, nilai produksi, dan jumlah nelayan ikan cakalang pada tahun 2023. *Cluster 1* terdiri atas provinsi-provinsi dengan volume dan nilai produksi yang rendah, namun memiliki jumlah nelayan yang relatif tinggi. Hal ini mengindikasikan bahwa produktivitas per nelayan di *cluster* ini masih rendah dan kegiatan perikananannya cenderung padat karya. *Cluster 2* mencakup provinsi-provinsi dengan volume dan nilai produksi pada tingkat menengah, namun memiliki jumlah nelayan paling sedikit dibandingkan *cluster*



lainnya. Karakteristik ini menunjukkan tingkat efisiensi atau produktivitas yang tinggi karena hasil produksi yang cukup besar dihasilkan oleh tenaga kerja yang lebih sedikit. Sementara itu, *cluster* 3 berisi provinsi-provinsi dengan volume produksi, nilai produksi, dan jumlah nelayan tertinggi. *Cluster* ini mencerminkan wilayah dengan sektor perikanan tangkap yang telah berkembang pesat, memiliki kapasitas produksi besar. Hasil pengelompokan ini dapat dijadikan dasar dalam merumuskan kebijakan pembangunan perikanan yang lebih merata dan sesuai dengan kondisi masing-masing wilayah.

REFERENSI

- [1] T. D. B. Diningrum, H. Triyono, And M. A. Jabbar, "Aspek Biologi Cakalang (Katsuwonus Pelamis, Linnaeus 1758) Di Sulawesi Tenggara," *Jurnal Penyuluhan Perikanan Dan Kelautan*, Vol. 13, No. 2, Pp. 139–147, Aug. 2019, Doi: 10.33378/Jppik.V13i2.195.
- [2] A. Sativa Aufania, E. Siti Rahayu, And S. Mulya Handayani, "Analisis Integrasi Pasar Ikan Cakalang (Katsuwonus Pelamis) Antar Nelayan Dan Pedagang Dki Jakarta," *Agrista*, Vol. 11, No. 3, Pp. 85–94, 2023.
- [3] A. Firlansyah, A. Akram, N. Risal, F. Adiba, And A. B. Kaswar, "Clustering Produksi Perikanan Budidaya Laut Berdasarkan Provinsi Menggunakan Algoritma K-Means." [Online]. Available: <https://ojs.unm.ac.id/jessi/index>
- [4] M. A. Nahdliyah, T. Widiharih, And A. Prahutama, "Metode K-Medoids Clustering Dengan Validasi Silhouette Index Dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota Di Jawa Tengah Tahun 2018)," Vol. 8, No. 2, Pp. 161–170, 2019, [Online]. Available: <http://ejournal3.undip.ac.id/index.php/gaussian>
- [5] I. Yanti, "Aplikasi Analisis Statistika Deskriptif Untuk Penataan Administrasi Dan Penyempurnaan Monografi Desa Slumbung Menuju Desa Mandiri Tertib Administrasi." [Online]. Available: <http://jiat.ub.ac.id>
- [6] F. Izzatul Muttaqin, Z. Rizky Ramadani, And R. Yotenka, "Implementasi Algoritma K-Means Dalam Analisis Cluster Usaha Pertanian Perorangan Di Provinsi D. I Yogyakarta Tahun 2023," *Emerging Statistics And Data Science Journal*, Vol. 3, No. 1, 2025.
- [7] M. Sriningsih, D. Hatidja, And J. D. Prang, "Penanganan Multikolinearitas Dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras Di Provinsi Sulut."
- [8] T. Rena Mayasari, F. Statistisi Pertama Bps Kabupaten Pesawaran Jl Ahmad Yani No, And G. Tatan, "Pengelompokkan Provinsi Berdasarkan Variabel Kesehatan Lingkungan Dan Pengaruhnya Terhadap Kemiskinan Di Indonesia Tahun 2018," 2020.
- [9] R. Putra Yusya, "Perancangan Ikat Pinggang Elektronik Untuk Tunanetra Menggunakan Mikrokontroller Dan Global Positioning System (Gps) Pada Smartphone Android," *Jurnal Teknik Elektro Itp*, Vol. 5, No. 2, 2016.
- [10] N. A. Maori, "Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering," *Jurnal Simetris*, Vol. 14, 2023.
- [11] T. Rahmawati, Y. Wilandari, And P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids," *Jurnal Gaussian*, Vol. 13, No. 1, Pp. 13–24, Aug. 2024, Doi: 10.14710/J.Gauss.13.1.13-24.
- [12] Kariyam, Abdurakhman, Subanar, H. Utami, And A. R. Effendie, "Block-Based K-Medoids Partitioning Method With Standardized Data To Improve Clustering Accuracy," *Mathematical Modelling Of Engineering Problems*, Vol. 9, No. 6, Pp. 1613–1621, 2022, Doi: 10.18280/Mmep.090622.
- [13] A. S. Sari, "Pengelompokkan Perguruan Tinggi Negeri Di Indonesia Berbasis Indikator Kinerja Utama Menggunakan Block-Based K-Medoids Partitioning Method," Universitas Islam Indonesia, Yogyakarta, 2023.
- [14] F. Alfiah, A. Almadayani, D. Al Farizi, And E. Widodo, "Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan Di Jawa Timur Tahun 2020," *Jurnal Ilmiah Sains*, Vol. 22, No. 1, P. 1, Dec. 2021, Doi: 10.35799/Jis.V22i1.35911.
- [15] A. Mahmudan, "Clustering Of District Or City In Central Java Based Covid-19 Case Using K-Means Clustering (Pengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Kasus Covid-19 Menggunakan K-Means Clustering)," Vol. 17, No. 1, Pp. 1–13, 2020, Doi: 10.20956/Jmsk.V%Vi%I.10727.
- [16] T. Soni Madhulatha, "An Overview On Clustering Methods," Vol. 2, No. 4, Pp. 719–725, 2012, [Online]. Available: www.iosrjen.org