

Analisis *Hierarchical* Dan *Non-Hierarchical Clustering* Untuk Pengelompokan Potensi Ekonomi Kelautan Indonesia 2021

Rinald Parulian Butar Butar¹

Program Studi D-IV Statistika Terapan Politeknik Statistika STIS¹
Jl. Otto Iskandardinata No.64C Jakarta 13330 Jakarta Timur Indonesia

¹212011442@stis.ac.id

Abstrak

Indonesia merupakan negara dengan wilayah laut yang sangat luas. Hal ini dibuktikan dengan hampir 70% wilayah Indonesia adalah laut dan sisanya daratan. Hal ini membuat pemerintah harus bersiap untuk memanfaatkan kekayaan laut yang dapat membangkitkan industri kelautan dan Perikanan. Sektor kelautan dan Perikanan masuk dalam Program Prioritas nasional yang diharapkan dapat meningkatkan pertumbuhan ekonomi nasional bagi negara. Namun untuk mencapai hal itu, Pemerintah membutuhkan kebijakan di sektor perikanan dan kelautan, tetapi minimnya pengelolaan Klasterisasi potensi perikanan menjadi hambatan bagi pemerintah. Tujuan penelitian yaitu pengelompokan potensi wilayah produksi perikanan di Indonesia tahun 2021. Penelitian ini membandingkan antara metode kluster hirarki yaitu *complete linkage*, *single linkage*, *average linkage* dan *Ward* dengan Metode kluster non-hirarki yaitu *k-means*, *k-medoids* dan *DBSCAN* (*Density-based Spatial Clustering of Application with Noise*). Penentuan metode terbaik kluster hirarki dan jumlah kluster yang paling optimum yang terdiri dari *average linkage*, *complete linkage*, *single linkage* dan *Ward* serta metode *Non - hierarchical clustering* yaitu *K-means* dan *k-medoids* adalah dengan *Davies-Bouldin Index* (DB), *Calinski-Harabasz* (CH), *Silhouette*, dan *Dunn Index*. Kemudian untuk metode *DBSCAN* adalah dengan membandingkan nilai *Silhouette*. Setelah ditentukan jumlah kluster terbaik, maka akan dilakukan evaluasi model dengan nilai *Internal Cluster Dispersion Rate*. Model *clustering* terbaik adalah *average linkage* dengan nilai *Internal Cluster Dispersion Rate* paling kecil dibandingkan model *cluster* lainnya. Maka dapat dilihat provinsi dengan potensi ekonomi kelautan yang paling tinggi dan yang paling rendah. Kluster *average linkage* menghasilkan 2 kluster. Kluster 1 beranggotakan 30 provinsi dengan potensi ekonomi kelautan yang perlu ditingkatkan. Kemudian kluster 2 yang beranggotakan 4 provinsi dengan potensi ekonomi kelautan yang tinggi.

Kata kunci: Ekonomi Kelautan, *Clustering*, *Hierarchical*, *Non-Hierarchical*, Perbandingan

Hierarchical and Non-Hierarchical Clustering Analysis for Classifying Indonesia's Sea Economic Potential in 2021

Abstract

Indonesia is a country with a very wide sea area. This is evidenced by almost 70% of Indonesia's territory is sea and land. This makes the government must be prepared to take advantage of marine wealth which can cause industrial and fisheries complications. Problematic sectors and fisheries are included in the national priority program which is expected to increase national economic growth for the country. However, to achieve this, the government needs policies in the fisheries and marine sector, but the lack of management of fishery potential clusters is an obstacle for the government. The research objective is to classify the potential of fishery production areas in Indonesia in 2021. This research compares the hierarchical cluster methods, namely complete linkage, single linkage, average linkage and Ward with the non-hierarchical cluster methods namely k-means, k-medoids and DBSCAN (*Density- Application-based Spatial Grouping with Noise*). The best hierarchical cluster operating method and the most optimum number of clusters consisting of average linkage, complete linkage, single linkage and ward as well as Non-hierarchical clustering methods namely K-means and k-medoids are with the Davies-Bouldin (DB) index, Calinski- Harabasz (CH), Silhouette, and Dunn Index. Then for the DBSCAN method is to compare the Silhouette values. After determining the best number of clusters, an evaluation of the model will be carried out with the Internal Cluster Dispersion Rate value. The best clustering model is the average linkage with the smallest Internal Cluster Dispersion Rate value compared to other cluster models. So it can be seen the provinces with the highest and lowest economic potential difficulties. The average linkage cluster produces 2 clusters. Cluster 1 consists of 30 provinces with potential economic chaos that needs to be improved. Then cluster 2 which consists of 4 provinces with high potential for economic difficulties.

I. PENDAHULUAN

Machine learning merupakan salah satu bagian dari data mining yang terus memunculkan inovasi dan berbagai teknik yang dapat menyelesaikan permasalahan data yang kompleks. Salah satu pendekatan di dalam *machine learning* adalah pendekatan *supervised* dan *unsupervised learning* dimana pendekatan tersebut merupakan yang paling umum dan sangat masif penggunaannya dibandingkan pendekatan lainnya. [1]

clustering adalah metode dari bagian pendekatan *unsupervised learning* yang digunakan untuk mengelompokkan sekumpulan data set sesuai dengan karakteristik dalam data atau bisa disebut pengelompokkan secara alami. Kemudian dalam *clustering* dapat dikelompokkan menjadi beberapa jenis yaitu Hierarchical *clustering* dan Non-hierarchical *clustering* yang penggunaannya memiliki kelebihan dan kekurangan. Kelemahan ini merupakan salah satu prinsip dalam data mining yaitu “No free lunch”. Artinya, tidak ada metode yang tidak mempunyai kelemahan. Namun, hal itu dapat ditekan dan diminimalisir dengan ketepatan metode pengelompokkan yang digunakan [2].

Metode dalam hierarchical *clustering* dibagi lagi menjadi subclass seperti *agglomerative* dan *divisive*. Contoh dari *agglomerative* misalnya *average linkage*, *complete linkage*, *single linkage*, *centroid linkage* dan *Ward*. Kemudian untuk metode dalam non-hierarchical *clustering* terdapat subclass yaitu *partition clustering*, *density based*, *grid base*, dll. Pendekatan *partition clustering* ini terdiri dari beberapa metode misalnya *k-means*, *K-medoids*, dll. Kemudian untuk pendekatan *density based* dapat digunakan metode *DBSCAN (Density-based Spatial Clustering of Application with Noise)* [3].

Indonesia merupakan negara dengan wilayah laut yang sangat luas. Hal ini dibuktikan dengan hampir 70% wilayah Indonesia adalah laut dan sisanya daratan. Hal ini membuat pemerintah harus bersiap untuk memanfaatkan kekayaan laut yang dapat membangkitkan industri kelautan dan Perikanan. Memanfaatkan kekayaan laut berarti menjadi bagian dari Upaya peningkatan di sektor ekonomi kelautan. Ekonomi kelautan bukan saja berbicara mengenai sektor perikanan baik tangkap maupun budidaya melainkan seluruh sektor yang berkaitan dengan perekonomian kelautan. Sektor yang berkaitan langsung dengan ekonomi kelautan selain perikanan yaitu industri pengolahan ikan, perkembangan bioteknologi laut, sumber daya alam yang berada di wilayah pulau pulau kecil. Pemanfaatan ini sangat penting bagi ekonomi Indonesia khususnya berbicara mengenai hal kesejahteraan bagi kehidupan nelayan [4].

Permasalahan yang dialami pemerintah adalah kurang mengoptimalkan produksi perikanan baik tangkap maupun budidaya yang dapat meningkatkan perekonomian negara yang bisa dilihat dari Produk Domestik Bruto di sektor perikanan. Kesulitan pemerintah ini juga ditambah bahwa potensi ekonomi kelautan di setiap provinsi yang ada di Indonesia berbeda beda. Namun tidak cukup hanya itu, kualitas infrastruktur dan alat penangkapan ikan juga perlu diperhatikan agar nelayan dapat menghasilkan bukan dalam kuantitas

banyaknya ikan yang diperoleh, tetapi juga memperhatikan produktivitas nelayan dalam menangkap ikan. Selain itu, kontribusi perikanan terhadap konsumsi masyarakat yang berbeda beda di setiap provinsi menjadi masalah yang kompleks.

Oleh karena itu, campur tangan pemerintah diperlukan untuk mengeluarkan kebijakan yang dapat meningkatkan pemanfaatan ekonomi kelautan. Perbedaan karakteristik kelautan dan potensi ekonominya di setiap provinsi yang ada di Indonesia dan belum adanya kajian yang mendalam tentang penyebaran potensi ekonomi kelautan baik dari segi produksi perikanan, kesejahteraan nelayan, konsumsi ikan dan peralatan nelayan akan dilakukan *clustering* berdasarkan provinsi yang ada di Indonesia. Algoritma yang digunakan adalah metode *hierarchical clustering* yaitu dengan *average linkage*, *complete linkage*, *single linkage* dan *ward* [5]. Kemudian akan dibandingkan dengan algoritma *non-hierarchical clustering* dengan *k-means*, *k-medoids*, dan *DBSCAN*. Selanjutnya akan dipilih jumlah kluster yang optimum dengan ukuran validitas *Calinski-Harabasz (CH)*, *Davies-Bouldin (DB)*, *Dunn* dan *koefisien Silhouette*. [6] Kemudian akan dievaluasi dengan Nilai Koefisien korelasi *cophenetic* dan *Internal Cluster Dispersion Rate* untuk dipilih sebagai sebagai metode kluster yang terbaik [7][8].

II. KAJIAN PUSTAKA

A. Hierarchical clustering

1) Metode complete linkage

Metode *complete linkage* merupakan bagian dari metode

Hierarchical clustering dimana pengelompokkannya dengan mencari jarak paling jauh diantara pengamatan suatu objek. Rumus yang digunakan untuk menentukan jarak antara (*uv*) dengan kluster lainnya adalah seperti berikut [5][9].

$$d_{(uv)k} = \max(d_{uk}, d_{vk}) \quad (1)$$

Keterangan:

d_{uk} : Jarak antara objek u dengan objek k

d_{vk} : Jarak antara objek v dengan objek k

$d_{(uv)k}$: Jarak paling maksimum antara objek *uv* dengan objek *k*

2) Metode average linkage

Metode *average linkage* merupakan bagian dari metode Hierarchical clustering dimana pengelompokkannya dengan mencari rata rata dari jarak diantara pengamatan suatu objek. Rumus yang digunakan untuk menentukan jarak antara (*u,v*) dengan kluster lainnya adalah seperti berikut [5].

$$d_{(uv)k} = \frac{\sum_x \sum_y d_{xy}}{n_{(uv)}n_k} \quad (2)$$

Keterangan :

d_{xy} : Jarak antara objek x pada cluster (*uv*) dan objek y cluster k

$n_{(uv)}$: Jumlah objek di cluster (*uv*)

n_k : Jumlah objek di cluster k

3) *Metode single linkage*

Metode *single linkage* merupakan bagian dari metode Hierarchical clustering dimana pengelompokkannya dengan mencari jarak paling dekat atau minimum diantara pengamatan suatu objek. Rumus yang digunakan untuk menentukan jarak antara (u, v) dengan klaster lainnya adalah seperti berikut [5][9].

$$d_{(uv)k} = \min (d_{uk}, d_{vk}) \quad (3)$$

Keterangan :

- d_{uk} : Jarak antara objek u dengan objek k
- d_{vk} : Jarak antara objek v dengan objek k
- $d_{(uv)k}$: Jarak paling minimum antara objek uv dengan objek k

4) *Metode Ward*

Metode Ward merupakan bagian dari metode Hierarchical clustering dimana pengelompokkannya berdasarkan nilai *Sum of Squares Error* (SSE) yang terkecil dan dijadikan satu klaster. Cara ini lebih memperhatikan keragaman dari suatu data yaitu mencari keragaman yang paling kecil yang berada dalam satu klaster. Rumus yang digunakan untuk mencari SSE adalah seperti berikut [5][11].

$$SSE = \sum_{j=1}^n (\sum_{i=1}^{n_j} Z_{uv}^2 - \frac{1}{n_j} (\sum_{i=1}^{n_j} Z_{uv})^2) \quad (4)$$

Dimana:

- Z_{uv} : nilai objek ke-u pada kelompok ke-v
- n : Jumlah pada kelompok tiap tahap
- n : Jumlah kelompok ke-u pada kelompok ke-v

B. *Non-Hierarchical clustering*

1) *K-means clustering*

K-Means clustering merupakan bagian dari metode hierarchical clustering yang termasuk pada pendekatan partisi. Tujuannya adalah untuk mencari jumlah jarak yang paling kecil diantara suatu titik data. Penentuan jumlah klaster pada metode ini harus diinisialisasikan terlebih dahulu sebelum dilakukan pengelompokkan. Berikut adalah Langkah langkahnya [11].

- a) Inisialisasikan jumlah klaster yang akan dibentuk
- b) Memutuskan titik yang akan dijadikan centroid secara random
- c) Membuat keputusan disetiap titik yang ada dengan centroid paling dekat, setelah ditentukan paling dekat akan terbentuk klaster yang sudah diinisialisasikan.
- d) cari varians dan ubah titik pusat baru ditiap klasternya
- d) lakukan iterasi titik pusat dari tiap cluster
- e) Sampai centroid tidak berubah

2) *K-medoids clustering*

K-Medoids clustering atau disebut dengan istilah *Partitioning Around Medoids* adalah pengembangan dari metode *k-means clustering*. Salah satu kelebihan dari metode ini adalah mampu mengatasi adanya data yang diindikasikan sebagai *outlier* yang dapat mempengaruhi distribusi data yang sebenarnya. Tahapan dalam menggunakan metode *k-medoids clustering* ini adalah sebagai berikut [12]:

- 1 Inisialisasikan jumlah klaster (k) yang akan dibentuk

- 2 Memilih random medoid di awal sebanyak jumlah klaster (k)
- 3 Mencari jarak dari setiap objek ke medoid awal, lalu mencari jarak terdekat dan jumlahkan total keseluruhannya. .
- 4 Iterasi dengan medoid yang baru
- 5 Cari simpangan data
- 6 Jika simpangan data kurang dari nol, maka lakukan Penggantian objek dengan membuat kumpulan k yang baru sebagai titik medoid
- 7 Terus lakukan perulangan dari langkah yang ketiga sampai ke lima dan berhenti saat tidak adanya perubahan pada anggota medoid.

3) *DBSCAN*

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah metode dalam non-hierarchical clustering yang masuk dalam pendekatan *density Based* dan bertujuan untuk menemukan tingkat kepadatan yang paling tinggi. Dalam DBSCAN tidak mencari jumlah klaster terlebih dahulu, namun menentukan berapa jarak antar titik dalam satu klaster yang dijadikan sebagai tetangga atau disebut dengan (Eps) dan berapa banyak titik minimal yang harus ada dalam suatu klaster atau disebut dengan (MinPts) [13]. Tahapannya adalah sebagai berikut:

- 1 Inisialisasikan nilai dari MinPts dan Eps
- 2 Pemilihan titik awal secara random
- 3 Menghitung jarak dari titik satu ke titik lain, dimana titiknya sudah dalam rentang jarak (Eps) yang sudah ditentukan diawal dengan menggunakan jarak Euclidean.
- 4 Maka akan terbentuk sebuah klaster dengan Eps dan MinPts yang diinisialisasikan di langkah 1.
- 5 Jika terdapat titik yang tidak memenuhi jarak (Eps) yang ditentukan tadi, maka akan dilanjutkan ke titik yang lain yang.
- 6 Kemudian lakukan terus perulangan langkah 3 sampai 4.

C. *Penentuan Jumlah klaster terbaik*

1) *Indeks Dunn (DN)*

Indeks validitas Dunn (DN) adalah salah satu ukuran validasi yang digunakan untuk mencari jumlah klaster yang optimum dari penentuan internal dan pengukuran yang dilakukan adalah seberapa besar varians pengelompokkan dan kekompakkan. Kekompakan satu klaster jika perbedaan dalam klaster yang sangat kecil yaitu dengan menghitung d_m seperti rumus dibawah ini [7][8].

$$d_m (c_k) = \max_{U_i, U_j \in c_k} d_e (U_i, U_j) \quad (5)$$

Untuk menghitung d_e yaitu dengan mencari jarak Euclidean.

Kemudian sebuah *cluster* dikatakan baik ketika antar *cluster* berjauhan dengan formula sebagai berikut

$$d(c_k, c_l) = \min_{U_i \in c_k, U_j \in c_l} d_e (U_i, U_j), \quad (6)$$

Kemudian untuk mencari Indeks Dunn adalah sebagai berikut :

$$DN = \min_{p=1, \dots, k} \left\{ \min_{q=i+1, \dots, k} \left(\frac{d_e(c_k, c_l)}{\max_{r=1, \dots, k} d_m(c_k)} \right) \right\} \quad (7)$$

2) Indeks *Davies – Bouldin* (DB)

Indeks validitas *Davies-Bouldin* (DB) adalah mencari rata-rata nilai tiap *point* pada himpunan data. Caranya adalah membandingkan antara total keseluruhan nilai compactness dengan jarak antara kedua titik centroid pada tiap klaster sebagai separation. Jumlah klaster terbaik ditunjukkan dengan nilai DB yang semakin kecil [7][8]. Misalkan terdapat suatu himpunan data dengan banyak k klaster, terdapat n_i *point* pada klaster X dan n_j pada klaster Y dengan C_x = titik pusat klaster X dan C_y = titik pusat klaster Y. kemudian jarak antara titik pusat X dan Y diberi simbol D_{xy} . S_x merupakan rata rata jarak dari tiap titik di klaster X ke titik pusat pada klaster yang satu (C_x) dan dengan klaster yang lain yaitu (C_y). S_y merupakan rata rata jarak dari tiap titik di klaster Y ke titik pusat pada klaster yang satu dengan klaster yang lain.

$$DB = \frac{1}{k} \sum_{p=1}^k R_x \quad (8)$$

$$R_x = \max R_{x,y}, X \neq Y, \quad (9)$$

$$R_{x,y} = \frac{(S_x + S_y)}{M_{xy}}, \quad (10)$$

$$S_x = \frac{1}{n_x} \sum_{i=1}^{n_x} d(U_i, c_x), \quad (11)$$

$$S_y = \frac{1}{n_y} \sum_{i=1}^{n_y} d(V_j, c_y), \quad (12)$$

$$M_{xy} = d(c_x, c_y) \quad (13)$$

3) Indeks *Calinski – Harabasz* (CH)

Indeks validitas *Calinski-Harabasz* (CH) adalah rasio nilai *Sum of Square between cluster* (SSB) sebagai separation dan nilai *Sum of Square within-cluster* (SSW) sebagai compactness kemudian dikali dengan faktor normalisasi, yaitu selisih jumlah data dengan jumlah klaster dibagi dengan jumlah klaster dikurang satu [7][8]. Jumlah klaster terbaik ditunjukkan dengan semakin besar nilai CH.

$$CH = \frac{\text{trace}(SSB)}{\text{trace}(SSW)} \times \frac{m-g}{g-1}, \quad (14)$$

$$SSW = \sum_{i=1}^g \sum_{x_i \in c_i} (x_i - \bar{x})(x_i - \bar{x})^T, \quad (15)$$

$$SSB = \sum_{i=1}^g M_i (x_i - \bar{x})(x_i - \bar{x})^T \quad (16)$$

Keterangan :

g : banyak klaster

m : banyak data

trace(SSB): dispersi antar klaster

trace(SSW): dispersi dalam klaster

4) *Silhouette*

Indeks validitas *silhouette* adalah salah satu ukuran validasi yang digunakan untuk mencari jumlah klaster yang optimum dari penentuan internal. Tujuannya adalah untuk melakukan evaluasi dengan melakukan perbandingan antara jarak rata rata dari suatu titik yang masuk ke dalam suatu klaster yang sama dengan suatu klaster yang lain Rumus yang digunakan pada *silhouette* adalah sebagai berikut [14]:

$$SILH = \frac{1}{N} \sum_{i=1}^N \left(\frac{u(i) - v(i)}{\max\{u(i), v(i)\}} \right) \quad (17)$$

Keterangan:

$u(i)$: jarak rata-rata antara suatu titik ke titik lain dalam klaster yang sama

$v(i)$: Jarak paling kecil antara suatu titik ke-i pada klaster yang lain.

Interpretasi dari indeks validitas *silhouette* adalah apabila nilainya mendekati 1 maka klaster yang terbentuk semakin bagus dan sebaliknya jika menjauhi 1 atau mendekati ke angka -1 maka semakin tidak bagus.

D. *Evaluasi klaster*

1) Koefisien korelasi *cophenetic*

Koefisien korelasi *cophenetic* adalah nilai yang menunjukkan korelasi antara elemen - elemen dari suatu matriks jarak Euclidian dengan elemen-elemen yang dihasilkan oleh matriks *cophenetic*. Rumus yang digunakan pada korelasi *cophenetic* adalah sebagai berikut[7][8].

$$cro = \frac{\sum_{u < v} (X_{uv} - \bar{x})(Y_{uv} - \bar{y})}{\sqrt{[\sum_{u < v} (X_{uv} - \bar{x})^2][\sum_{u < v} (Y_{uv} - \bar{y})^2]}} \quad (18)$$

keterangan:

cro : koefisien korelasi *cophenetic*

X_{uv} : jarak asli (jarak Euclidean) antara objek u dan v dari X

\bar{x} dan \bar{y} : rata-rata nilai x dan y

Y_{uv} : jarak *cophenetic* antara objek u dan v dari Y

2) *Internal Cluster Dispersion Rate* (Icdrate)

Ukuran evaluasi lainnya yang digunakan adalah *Internal Cluster Dispersion Rate*. Evaluasi ini mencari nilai rata – rata dari disperse sebaran internal klaster terhadap partisi keseluruhan. *Icdrate* (*internal cluster dispersion*) menggambarkan tingkat disperse atau perbedaan dalam klaster. Klaster yang terbaik dilihat dari nilai *icdrate* yang paling kecil. Rumus yang digunakan pada *Internal Cluster Dispersion Rate* adalah sebagai berikut [7][8]:

$$icdr = 1 - \frac{SSB}{SST} = 1 - r^2 \quad (19)$$

Keterangan :

SSB : *Sum Square Between*

SST : *Sum Square Total* atau Total dari jumlah kuadrat jarak sampel terhadap rata-rata keseluruhan

Icdr : *Internal Cluster Dispersion Rate*

III. METODE PENELITIAN

Penelitian ini mengaplikasikan proses data mining yaitu :

1. Pengumpulan data

Pengumpulan data dilakukan dengan mencari variabel variabel yang berkaitan dengan potensi ekonomi kelautan berdasarkan situs dari kementerian perikanan dan kelautan dan website resmi Badan Pusat Statistik.

2. Preprocessing

Teknik *Preprocessing* data yang digunakan adalah membersihkan data dan melakukan transformasi data dengan melakukan konversi data asli menjadi *z-score*.

3. Penerapan model

Model data mining yang diterapkan pada penelitian ini data hasil *preprocessing* dengan *complete linkage*,

average linkage, single linkage, ward, k-means, k-medoids dan DBSCAN

4. Validasi jumlah kluster terbaik
Untuk menentukan jumlah *cluster* terbaik digunakan indeks validitas Dunn (DN), Davies – Bouldin (DB), Calinski – Harabasz (CH) dan Silhouette
5. Evaluasi kluster terbaik
Evaluasi kluster terbaik dilakukan dengan membandingkan Koefisien korelasi cophenetic ditiap kluster yang terpilih dari validasi jumlah kluster terbaik, dan dengan *Internal Cluster Dispersion Rate* (Icdrate)
6. Hasil Clustering
Setelah dilakukan evaluasi akan diperoleh hasil *clustering* terbaik potensi ekonomi kelautan di Indonesia.

Penelitian ini menggunakan data sekunder yang diperoleh dari Website Kementerian Kelautan dan Perikanan, Publikasi Kelautan dan Perikanan dalam angka tahun 2021[15] dan website Badan Pusat Statistik [16]. Atribut yang digunakan sebanyak 9 variabel pada sektor Ekonomi Kelautan dari ke-34 provinsi yang ada di Indonesia. Berikut adalah rincian dari 9 variabel yang digunakan pada Tabel 1 dibawah ini.

TABEL I
VARIABEL PENELITIAN SEKTOR EKONOMI KELAUTAN DI INDONESIA TAHUN 2021

kode	Atribut	Definisi atribut	Satuan
PPT	Produksi Perikanan Tangkap	Hasil tangkapan ikan di laut maupun perairan umum di masing masing provinsi	ton
PPB	Produksi perikanan Budidaya pembesaran	Hasil tangkapan ikan dari Budidaya ikan baik budidaya pembenihan, budidaya air payau, budidaya air laut, dan budidaya air tawar	ton
AKI	Angka Konsumsi Ikan	Konsumsi ikan oleh masyarakat Indonesia yang diukur dalam satuan kilogram per kapita per tahun	Kg/Perkapita
PDRBI	PDRB ADHK sektor Perikanan	Produk domestik regional Bruto atas dasar Harga Konstan dibidang sektor perikanan	Miliar rupiah
RTP	Jumlah Rumah Tangga Perikanan	Jumlah rumah tangga yang melakukan kegiatan budidaya ikan baik di Laut maupun perairan umum	unit

	Tangkap		
KL	Jumlah Kapal Perikanan Laut Kapal Motor 10 - 30 GT	Jumlah kapal Perikanan laut berukuran 10 sampai 30 Gross ton	unit
NTN	Nilai Tukar Nelayan	Rasio antara harga yang diterima nelayan dengan harga yang dibayarkan nelayan	%
NTPI	Nilai Tukar pembudidaya Ikan	Rasio kemampuan tukar hasil ikan budidaya dengan barang/jasa untuk produksi dan konsumsi Rumah tangga	%
UPI	Jumlah Unit Pengolahan Ikan	lokasi dan sarana untuk melakukan kegiatan pengolahan ikan	unit

IV. HASIL DAN PEMBAHASAN

Hasil pengolahan statistik dalam penelitian ini melibatkan penggunaan RStudio versi 4.2.0 sebagai program bantuannya. Berikut ini adalah output dan pembahasan yang diperoleh dari pengolahan data menggunakan RStudio.

A. Analisis Deskriptif

Analisis ini bertujuan untuk mendeskripsikan ke-9 atribut potensi ekonomi kelautan yang ada di Indonesia berdasarkan ke-34 provinsi. Berikut ini adalah rincian dari statistik deskriptif yang terdiri dari nilai minimum, nilai maksimum, mean dan median dari dataset Potensi ekonomi kelautan Indonesia tahun 2021. Analisis deskriptif diperlihatkan pada Tabel 2 dibawah ini.

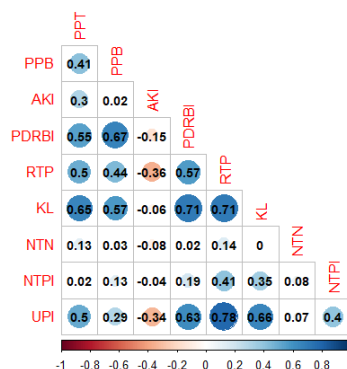
TABEL II
STATISTIK DESKRIPTIF DATA POTENSI EKONOMI KELAUTAN TAHUN 2021

Variabel	Minimum	Maximum	Mean	median
Produksi Perikanan Tangkap	7145	547463	212486	193044
Produksi perikanan Budidaya pembesaran	2370	4082792	430833	146550
Angka Konsumsi Ikan	34.82	77.49	55.37	56.12
PDRB ADHK sektor	346	33908	6938	4772

Perikanan				
Jumlah Rumah Tangga Perikanan Tangkap	6174	386516	85012	68529
Jumlah Kapal Perikanan Laut Kapal Motor 10 - 30 GT	8	4814	974	475.5
Nilai Tukar Nelayan	92.55	115.84	104.50	104.58
Nilai Tukar pembudidaya Ikan	91.33	110.21	99.64	100.56
Jumlah Unit Pengolahan Ikan	142.0	10624	1705.5	862.0

Berdasarkan sajian Tabel II, rata-rata Produksi Perikanan Tangkap dan Produksi Perikanan Budidaya Pembesaran di Indonesia tahun 2021 adalah masing masing sebesar 212486 ton dan 430833 ton. Produksi perikanan budidaya pembesaran menjadi produksi yang paling besar dibandingkan produksi perikanan tangkap. Selanjutnya, rata-rata angka konsumsi ikan adalah sebesar 55.37 Kg per kapita sepanjang tahun 2021. Selain itu, rata rata Produk Domestik regional Bruto ADHK untuk sektor Perikanan di Indonesia tahun 2021 adalah sebesar 6938 Miliar rupiah, dengan nilai tertinggi mencapai 33908 miliar rupiah. Rata rata Jumlah Rumah Tangga perikanan Tangkap Indonesia Tahun 2021 adalah sebanyak 85012 rumah tangga dengan jumlah paling sedikit adalah sebanyak 6174 Rumah Tangga dan terbanyak sebesar 386516. Kemudian rata – rata Jumlah Kapal Perikanan Laut Kapal Motor 10 - 30 GT di Indonesia sebanyak 974 unit dan paling banyak sebesar 4814 unit. Nilai Tukar Nelayan dan Nilai Tukar pembudidaya ikan masing masing memiliki rata rata sebesar 104.50% dan 99.64%. kemudian untuk jumlah unit pengolahan ikan di Indonesia adalah sebesar 1705.5 dengan jumlah paling sedikit hanya 142 unit dan paling banyak sebesar 10642 unit.

B. Pendeteksian Multikolinearitas antar variable Independen

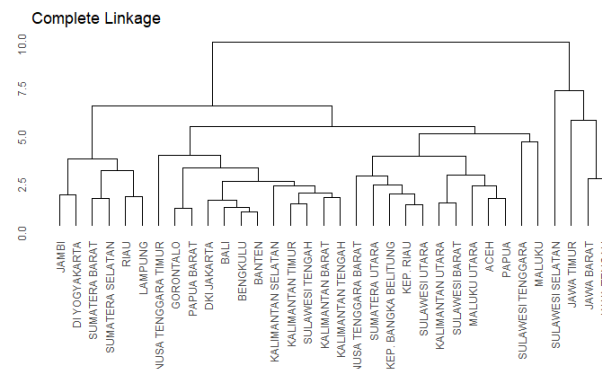


Gambar 1. Matriks korelasi antar variable

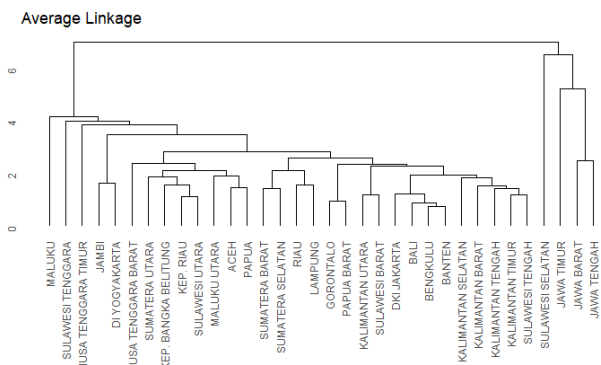
Dalam analisis matriks korelasi tidak ditemukan bahwa variabel-variabel dalam Ekonomi kelautan tidak memiliki korelasi positif yang cukup tinggi satu sama lain. Lebih lanjut, dalam mengidentifikasi gejala multikolinearitas tidak adanya korelasi yang melebihi 0,8 antara variabel-variabel independen [18]. Dengan demikian, dapat disimpulkan bahwa tidak terdapat masalah multikolinearitas, sehingga asumsi untuk analisis pengelompokan (clustering analysis) dapat dilakukan lebih lanjut.

C. Hierarchical Clustering

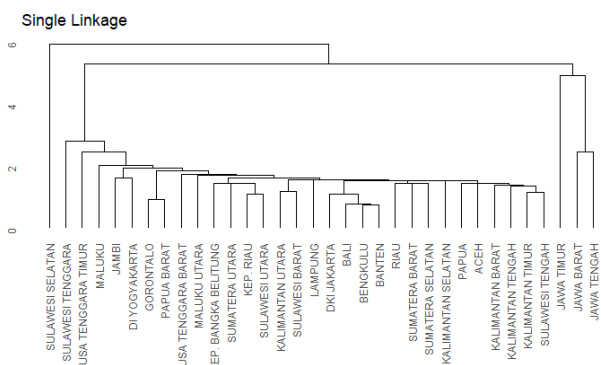
Pada algoritma Hierarchical Clustering digunakan empat metode, yaitu *single linkage*, *average linkage*, *complete linkage* dan *Ward*. Pada metode kluster hierarkis, jumlah klasternya tidak ditentukan diawal melainkan melihat dendogram yang dihasilkan dari metode diatas, kemudian setelah itu akan ditentukan jumlah klister yang akan dibuat berdasarkan kalster yang terbentuk dari dendogram tersebut. Berikut ini adalah dendogram yang dihasilkan oleh metode tersebut.



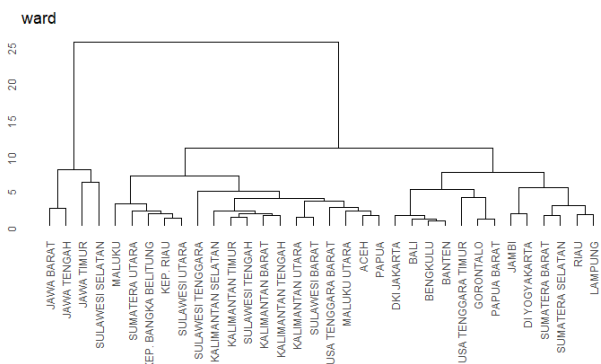
Gambar 2. Dendogram Complete Linkage



Gambar 3. Dendrogram Average Linkage



Gambar 4. Dendrogram Single Linkage



Gambar 5. Dendrogram Ward

Berdasarkan dendrogram dari gambar pada 2, 3 4 dan 5 diatas dapat dilihat bahwa terdapat perbedaan cluster dendrogram yang dihasilkan pada data Potensi Ekonomi Kelautan di Indonesia tahun 2021. Namun, secara keseluruhan dapat dilakukan pengelompokkan menjadi 2, 3, 4, dan 5 kluster. Dalam penentuan Jumlah kluster yang optimum dan metode yang terbaik dilakukan perhitungan nilai indeks validitas menggunakan Calinski-Harabasz (CH), Davies-Bouldin (DB), Dunn dan koefisien Silhouette. Selanjutnya perhitungan indeks validitas tersebut dilakukan pada metode *single linkage*, *average linkage*, *complete linkage* dan *Ward* dari jumlah kluster 2 hingga 5 kluster. Informasi lebih lanjut mengenai nilai-nilai indeks validitas terdapat pada Tabel III dibawah ini.

TABEL III

PERBANDINGAN HASIL INDEKS VALIDITAS DB, CH, DUNN DAN SILHOUETTE ALGORITMA HIERARCHICAL

Metode	Jumlah kluster	DB	CH	Dunn	Silhouette
complete	2	0.9602	20.819	0.7578	0.5424
	3	0.6598	14.537	0.8518	0.5354
	4	1.1929	12.339	0.2801	0.1564
	5	0.936	11.35	0.2972	0.1605
average	2	0.9602	20.819	0.7578	0.5424
	3	0.6598	14.537	0.8518	0.5354
	4	0.428	11.825	0.7898	0.5109
	5	0.4418	10.295	0.326	0.2796
single	2	0.4193	5.3531	0.6221	0.4798
	3	0.6598	14.537	0.8518	0.5354
	4	0.428	11.825	0.7898	0.5109
	5	0.4543	10.092	0.4912	0.2817
ward	2	0.9602	20.819	0.7578	0.5424
	3	1.6159	14.683	0.2189	0.1734
	4	1.3916	13.468	0.2573	0.1843
	5	1.247	12.34	0.2573	0.1837

Pada keempat indeks validitas dari table diatas untuk menentukan jumlah kluster yang optimum dan terbaik dapat dibandingkan dari Nilai Calinski-Harabasz (CH), Dunn dan koefisien Silhouette yang paling besar dan indeks Davies-Bouldin (DB) paling kecil pada masing masing metode. Hasil diatas didapatkan metode *complete linkage* menghasilkan jumlah kluster yang optimum adalah 2 dan 3. Berikutnya metode *average linkage* menghasilkan jumlah kluster yang optimum adalah 2. Selanjutnya metode *single linkage* menghasilkan jumlah kluster yang optimum adalah 3. Selanjutnya untuk metode *ward* menghasilkan jumlah kluster yang optimum adalah 2.

D. Non-Hierarchical Clustering

Non – hierarchical clustering yang digunakan dalam penelitian ini adalah pendekatan partisi (*partition based clustering*) dan Pendekatan *Density Based*. Pendekatan Partisi yang digunakan adalah *k-means* dan *k-medoids*. Kemudian Pendekatan *Density based* yang digunakan adalah DBSCAN (*Density-based Spatial Clustering of Application with Noise*).

1) Kluster *K-means* dan *k-medoids*

Pendekatan Kluster *k-means* dan *k-medoids* merupakan salah satu pendekatan partisi dalam *Non-Hierarchical Clustering*. Dalam kluster *k-means* jumlah kluster nya ditentukan di awal. Jumlah kluster yang akan di validasi adalah 2, 3, 4 dan 5.

TABEL IV

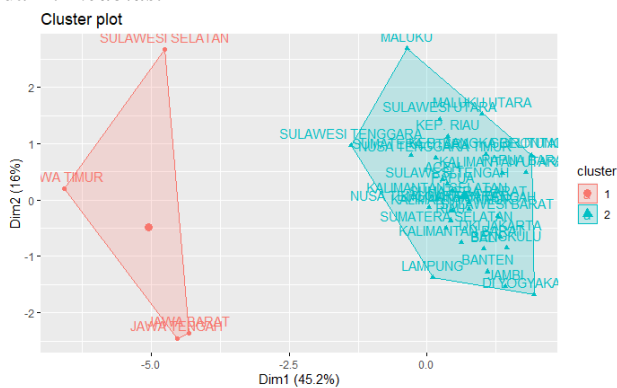
HASIL INDEKS VALIDITAS DB, CH, DUNN DAN SILHOUETTE ALGORITMA NON-HIERARCHICA (*K-MEANS* DAN *K-MEDOIDS*)

Metode	Jumlah kluster	DB	CH	Dunn	Silhouette
--------	----------------	----	----	------	------------

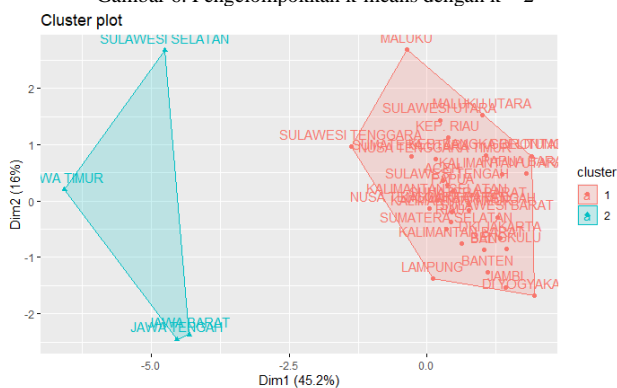
k-means	2	0.9602	20.819	0.7578	0.5424
	3	1.7074	14.184	0.8518	0.5354
	4	1.2537	12.465	0.7898	0.5109
	5	1.2541	13.062	0.2749	0.214
k-medoids	2	0.9602	20.819	0.7578	0.5424
	3	1.5568	13.117	0.1912	0.2252
	4	1.1459	14.249	0.2606	0.2093
	5	1.2897	12.317	0.2551	0.1505

Berdasarkan hasil validitas diatas dapat dilihat bahwa metode *k-means* menghasilkan jumlah kluster yang optimum adalah 2 karena Nilai DB paling kecil dan Nilai C, Dunn dan Silhouette paling besar. Kemudian untuk metode *k-medoids* menghasilkan jumlah kluster yang optimum adalah 2.

Berikut adalah hasil *clustering* dengan Metode *k-means* dan *k-medoids*.



Gambar 6. Pengelompokan k-means dengan k = 2

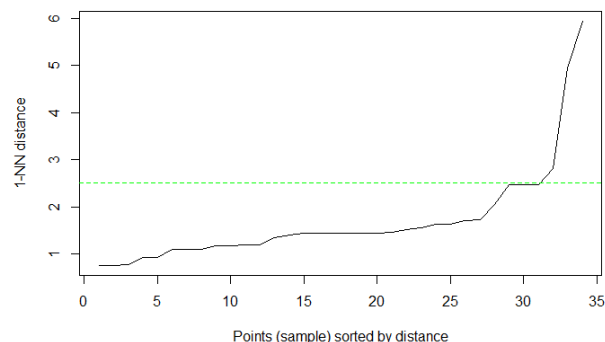


Gambar 7. Pengelompokan k-medoids dengan k = 2

2) *Kluster DBSCAN (Density-based Spatial Clustering of Application with Noise)*

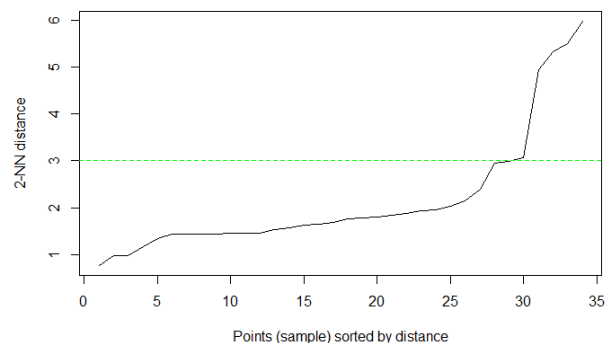
Dalam metode DBSCAN, untuk melakukan proses clustering, DBSCAN membutuhkan dua parameter input yaitu Epsilon (Eps) dan Minimum Points (MinPts). Epsilon merupakan jarak maksimal yang ditentukan antar dua data dalam satu kluster sedangkan Minimum Points adalah minimal jumlah data dalam satu kluster yang dibentuk. Penentuan Epsilon dan Minimum points berdasarkan nilai rata rata dari jarak dari setiap data ke tetangga paling dekat. Penentuan epsilon menggunakan percobaan yaitu dengan epsilon 2.5 hingga 3.05

Kemudian untuk MinPoints yang digunakan adalah 1, 2 dan 3. Berikut adalah rincian plot yang digunakan untuk menentukan nilai optimal dari epsilon dan minpoints yang sudah ditentukan.



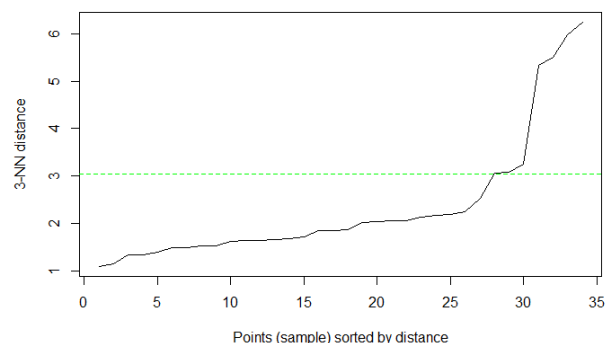
Gambar 8. Plot kNN dengan Epsilon 2,5 dan K = 1

Berdasarkan Plot diatas dapat dilihat dengan nilai k=1, diperoleh jarak yang optimum yaitu 2,5. Hal ini dikarenakan posisi siku pada plot diatas berada di angka 2,5



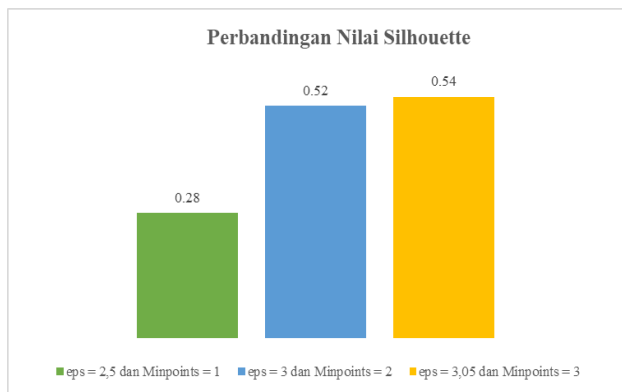
Gambar 9. Plot kNN dengan Epsilon 3 dan K = 2

Berdasarkan Plot diatas dapat dilihat dengan nilai k=2, diperoleh jarak yang optimum yaitu 3. Hal ini dikarenakan posisi siku pada plot diatas berada di angka 3



Gambar 8. Plot kNN dengan Epsilon 3,05 dan K = 3

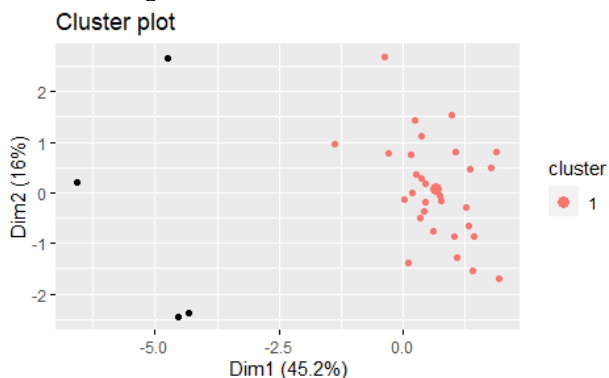
Berdasarkan Plot diatas dapat dilihat dengan nilai k=3, diperoleh jarak yang optimum yaitu 3,05. Hal ini dikarenakan posisi siku pada plot diatas berada di angka 3,5. Selanjutnya akan dilakukan validasi dengan mencari nilai Silhouette paling besar dari percobaan nilai Epsilon dan MinPoints. Berikut adalah Grafik dari Perbandingan Nilai Silhouette DBSCAN.



Gambar 10. Perbandingan Nilai Silhouette DBSCAN

Berdasarkan Hasil Perbandingan Indeks Validitas antara DBSCAN epsilon 2,5 hingga 3,05 Kemudian untuk MinPoints yang digunakan adalah 1, 2 dan 3 dengan dengan Nilai Silhouette paling besar adalah 0,54. Dengan demikian, dalam *Non-Hierarchical Clustering* akan dipilih DBSCAN dengan epsilon sebesar 3,05 dan MinimumPoints sebesar 3.

Berikut adalah hasil *clustering* dengan Metode DBSCAN dengan



Gambar 11. Visualisasi DBSCAN Eps = 3,05 dan MinPts =3

3) *Evaluasi Model Terbaik*

Metode terbaik yang akan dipilih untuk analisis kluster Hierarchical Clustering adalah metode dengan nilai koefisien Korelasi Cophenetic yang mendekati 1. Setelah melihat hasil koefisien korelasi cophenetic, kemudian akan memilih metode yang memiliki koefisien korelasi yang paling dekat dengan 1 sebagai metode terbaik.

Evaluasi Model terbaik dari Hierarchical Clustering dengan *Cophenetic Correlation Coefficient*

TABEL V

HASIL EVALUASI DARI ALGORITMA HIERARCHICAL CLUSTERING

Metode Clustering	Nilai Koefisien korelasi cophenetic
<i>complete linkage</i>	0.845
<i>average linkage</i>	0.937
<i>single linkage</i>	0.915
<i>ward</i>	0.871

Berdasarkan Nilai Koefisien korelasi cophenetic maka didapatkan hasil bahwa metode Hierarchical yang terbaik adalah dengan menggunakan *average linkage* karena nilai *Koefisien korelasi cophenetic* yang paling besar dan

mendekati 1 yaitu sebesar 0,937. Maka akan dipilih *average linkage* menghasilkan jumlah kluster yang optimum adalah 2 sesuai dengan kriteria terbaik yang sebelumnya sudah dilakukan. Selanjutnya adalah membandingkan nilai *internal Cluster dispersion Rate (Icdrate)*.

TABEL VI

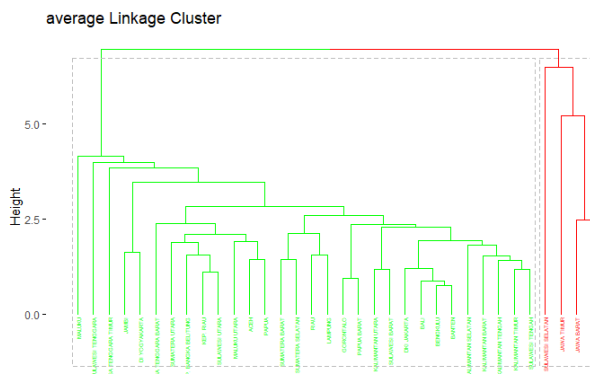
HASIL EVALUASI DARI ALGORITMA HIERARCHICAL DAN NON-HIERARCHICAL CLUSTERING

Metode Clustering	<i>Internal Cluster Dispersion Rate (Icdrate)</i>
<i>Average Linkage (k=2)</i>	0
<i>k-means</i>	37.805
<i>k-medoids</i>	36.852
DBSCAN	36.852

Berdasarkan Hasil Evaluasi Clustering dari Hierarchical *clustering* yang terpilih yaitu *average linkage* dengan jumlah kluster sebanyak 2 dan dibandingkan dengan *k-means*, *k-medoids* dan DBSCAN, nilai *Internal Cluster Dispersion Rate* yang paling rendah adalah metode *average linkage* dengan Kluster sebanyak 2. Dengan demikian, Metode terbaik yang digunakan untuk menentukan Clustering Potensi Ekonomi Kelautan adalah dengan *average linkage* dengan Kluster sebanyak 2.

4) *Metode terbaik (Average Linkage Clustering)*

Pada Pengelompokan *average Linkage Clustering* dihasilkan 2 kluster dengan Hasil Visualisasi dendrogram seperti berikut



Gambar 10. Dendrogram Average Linkage dengan 2 kluster

Berdasarkan dendrogram diatas dapat dilihat bahwa dendrogram yang berwarna hijau menunjukkan *cluster 1* dan dendrogram warna merah menunjukkan *cluster 2*. Kemudian setelah dilakukan Visualisasi akan ditampilkan nilai rata rata dari variable di tiap jumlah klasternya untuk kluster 1 dan 2. Berikut adalah hasil dari rata rata nilai dari setiap variable pada kluster 1 dan 2.

TABEL VII

NILAI RATA RATA VARIABEL DI TIAP KLASTER

Klaster	PPT	PPB	AKI	PDRBI	RTP
1	-0.1838	-0.2398	0.0988	-0.2638	-0.3092
2	1.3789	1.7986	-0.741	1.9785	2.3193

Klaster	KL	NTN	NTPI	UPI
1	-0.2749	-0.0488	-0.1682	-0.2941
2	2.0618	0.3663	1.2618	2.2057

Dari Hasil Pembentukan Klaster dengan Metode *average Linkage* dapat dilihat bahwa *cluster 2* menghasilkan nilai Rata rata yang paling tinggi dibandingkan dengan rata rata di *cluster 2*.

Selain dari nilai rata rata variable di tiap klasternya yang didapatkan, berikutnya juga akan di tampilkan rincian dari masing masing anggota tiap *clusternya* dari ke-34 Provinsi menjadi 2 *cluster*.

TABEL VIII
HASIL CLUSTERING AVERAGE LINKAGE

Cluster	Jumlah Anggota	Anggota
1	30	Maluku, Sulawesi Tenggara, Nusa Tenggara Timur, Jambi, DI.Yogyakarta, Nusa Tenggara Barat, Sumatera Utara, Kep. Bangka Belitung, Kep. Riau, Sulawesi Utara, Maluku Utara, Aceh, Papua, Sumatera Barat, Sumatera Selatan, Riau, Lampung, Gorontalo, Papua Barat, Kalimantan Utara, Sulawesi Barat, DKI Jakarta, Bali, Bengkulu, Banten, Kalimantan Selatan, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur dan Sulawesi Tengah.
2	4	Sulawesi Selatan, Jawa Timur, Jawa Barat dan Jawa Tengah.

Cluster 1 menghasilkan jumlah anggota paling banyak yaitu 30 anggota provinsi, sedangkan *Cluster 2* menghasilkan 4 anggota provinsi saja. Selain itu, dapat dilihat bahwa ternyata hasil produksi perikanan tangkap dan perikanan budidaya pembesaran paling dominan berada di *cluster 2* yaitu berada di Provinsi Sulawesi Selatan, Jawa Timur, Jawa Barat dan Jawa Tengah. Hal ini sesuai dengan hasil penelitian yang dilakukan oleh Sulistiyawan, dkk (2021) yang menggunakan data sektor perikanan tahun 2016 bahwa wilayah dengan sektor perikanan yang tinggi terdapat di Jawa Barat, Jawa Tengah dan Jawa Timur [17]. Namun, Hal yang berbeda ditunjukkan pada angka konsumsi ikan, hasil pengelompokkan menunjukkan bahwa *cluster 1* menghasilkan angka konsumsi ikan yang tinggi dibandingkan *cluster 2*. Hal yang menarik adalah Angka konsumsi ikan di *cluster 1* lebih tinggi dari pada di *cluster 2*, padahal produksi perikanan tangkap dan budidaya di *cluster 1* lebih rendah dibandingkan *cluster 2*. PDRB di sektor perikanan menunjukkan hasil yang sesuai bahwa tingginya produksi perikanan tangkap dan budidaya di *cluster 2* membuat PDRB di sektor perikanan lebih tinggi di bandingkan di *cluster 1*. Begitu juga dengan Provinsi dimana jumlah anggota rumah tangga perikanan tangkap yang lebih tinggi berada di anggota *cluster 2* dan Jumlah kapal perikanan laut kapal 10 – 30 GT membuat hasil produksi Perikanan tangkap lebih tinggi dibandingkan dengan provinsi yang berada di *cluster 1*. Selanjutnya, provinsi yang termasuk di anggota *cluster 2* menunjukkan tingkat kesejahteraan nelayan yang lebih baik dibandingkan provinsi yang masuk ke dalam *cluster 1*. Hal ini ditunjukkan dari nilai tukar nelayan yang lebih tinggi di *cluster 2*. Kemudian untuk nilai tukar pembudidaya ikan juga lebih tinggi di *cluster 2* dan juga jumlah unit pengolahan yang lebih banyak berada di *cluster 2*.

V. KESIMPULAN

Berdasarkan Hasil Perbandingan dengan metode *clustering* Hierarki dan Non-Hierarki menunjukkan bahwa pengelompokkan provinsi di Indonesia berdasarkan potensi Ekonomi kelautan tahun 2021 menghasilkan: validasi jumlah klaster yang terbaik di metode Hierarchical Clustering adalah dengan *average linkage* dengan 2 kelompok. Kemudian dari metode Non-hierarchical *clustering* didapatkan metode *k-means* dan *k-medoids* menghasilkan indeks validasi yang sama baik dengan menggunakan validitas Dunn (DN), Davies – Bouldin (DB), Calinski – Harabasz (CH) dan Silhouette. Pendekatan *density-based* yang terpilih adalah DBSCAN dengan Epsilon 3,05 dan MinPts sebesar 3. Selanjutnya dilakukan evaluasi untuk mendapatkan pengelompokkan terbaik dengan menggunakan Internal Cluster Dispersion Rate, didapatkan hasil Internal Cluster Dispersion Rate yang terkecil adalah *average linkage* dengan jumlah klaster sebanyak 2. Kesimpulannya adalah Pengelompokkan yang paling tepat digunakan untuk mengelompokkan potensi ekonomi kelautan di Indonesia

tahun 2021 adalah dengan metode *average linkage*. Berikut hasil rincian dari klater yang terbentuk.

- 1) Klaster 1 beranggotakan 30 provinsi dengan karakteristik variabel produksi perikanan baik tangkap maupun budidaya yang rendah. Hal ini dibuktikan dengan hasil rata rata klaster 1 lebih rendah dari klaster 2. Begitu juga dengan variabel lainnya seperti PDRB sektor perikanan, Jumlah rumah tangga perikanan tangkap, Jumlah Kapal Perikanan Laut Kapal Motor 10 - 30 GT, Nilai tukar nelayan, Nilai tukar pembudidaya ikan dan Jumlah unit pengolahan yang perlu diperhatikan oleh pemerintah. Namun lain halnya dengan angka konsumsi ikan yang lebih tinggi di bandingkan dengan klaster 2. Rendahnya hasil produksi perikanan di klaster 1 karena kurangnya pemanfaatan teknologi untuk menangkap ikan yang berdampak pada kesejahteraan nelayan yang sangat kurang. Hal ini bisa dilihat dari Nilai tukar nelayan yang lebih rendah dibandingkan klaster 2.
- 2) Klaster 2 beranggotakan 4 provinsi yaitu Sulawesi Selatan, Jawa Timur, Jawa Barat dan Jawa Tengah dengan variabel produksi perikanan baik tangkap dan budidaya ikan yang tinggi dan Angka konsumsi ikan yang rendah.

DAFTAR PUSTAKA

- [1] Fajriana, F. (2021). *Analisis Algoritma K-Medoids pada Sistem Klasterisasi Produksi Perikanan Tangkap Kabupaten Aceh Utara*. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 7(2), 263. <https://doi.org/10.26418/jp.v7i2.47795>
- [2] Gulagiz, F. K., & Suhap, S. (2017). *Comparison of Hierarchical and Non-Hierarchical Clustering Algorithms*. *International Journal of Computer Engineering and Information Technology*, 9(1), 6–14. www.ijceit.org
- [3] Khabaza, T. (2010). *9 Laws of Data Mining*. September, 1–10. http://khabaza.codimension.net/index_files/9laws.htm
- [4] "Potensi Perikanan Indonesia". Wantimpres.go.id. <https://wantimpres.go.id/id/2017/04/potensi-perikanan-indonesia/>
- [5] Purwaningsih, T., & Stat, S. (2017). *Perbandingan Hasil Pengelompokan antara Metode Average Linkage, Ward, Complete Linkage, dan Single Linkage (Studi Kasus: Indikator Kesehatan Indonesia)*. [skripsi] Yogyakarta : Universitas Islam Indonesia.
- [6] Khairati, A. F., Adlina, A. ., Hertono, G. ., & Handari, B. . (2019). *Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA*. PRISMA, Prosiding Seminar Nasional Matematika, 2, 161–170.
- [7] Putri, M. M. (2015). *Pengelompokan kabupaten/kota di Provinsi Jawa Timur berdasarkan Indikator Kesehatan Masyarakat menggunakan Metode Kohonen SOM dan k-means*. [skripsi] Surabaya : Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember.
- [8] Iis, Yahya, I., Wibawa, G. N. A., & Laome, L. (2022). *Penggunaan Korelasi Cophenetic untuk Pemilihan Metode Cluster Berhierarki pada Mengelompokkan Kabupaten/Kota Berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020*. Prosiding Seminar Nasional Sains Dan Terapan (SINTA) VI 2022, April, 1–16.
- [9] Eka, S., Setiawan, D., Mubarak, R., Ernawati, & Otok, B. W. (2017). *Analisis Cluster pada Kabupaten di Papua Berdasarkan*. May, 0–14.
- [10] Paramadina, M., Sudarmin, S., & Aidid, M. K. (2019). *Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan)*. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 1(2), 22. <https://doi.org/10.35580/variansiunm9357>.
- [11] Tempola, F., & Assagaf, A. F. (2018). *Clustering of Potency of Shrimp In Indonesia With K-Means Algorithm And Validation of Davies-Bouldin Index*. 1(lcst), 730–733. <https://doi.org/10.2991/icst-18.2018.148>
- [12] Nurkholik, D. (2022). *ANALISIS K-MEDOIDS CLUSTERING METODE ELBOW PADA KASUS COVID-19 DI PROVINSI DKI JAKARTA*. [Skripsi] Malang : Universitas Islam Negeri Maulana Malik Ibrahim.
- [13] Adha, R., Nurhaliza, N., Sholeha, U., & Mustakim, M. (2021). *Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia*. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 18(2), 206–211.
- [14] Rousseeuw, P. J. (1987). *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- [15] [KKP] Kementerian Kelautan dan Perikanan (2021). https://statistik.kkp.go.id/home.php?m=prod_ikan_prov&i=2# (diakses 20 juni 2023).
- [16] Badan Pusat Statistik (2023). <https://www.bps.go.id/subject/56/perikanan.html#subjekViewTab3> (diakses 20 Juni 2023).
- [17] Sulistiyawan, E., Hapsery, A., & Arifahanum, L. J. A. (2021). *PERBANDINGAN METODE OPTIMASI UNTUK PENGELOMPOKAN PROVINSI BERDASARKAN SEKTOR PERIKANAN DI INDONESIA (Studi Kasus Dinas Kelautan dan Perikanan Indonesia)*. *Jurnal Gaussiana*, 10(1), 76–84. <https://doi.org/10.14710/j.gauss.v10i1.30936>
- [18] Faujia, R. A., Setianingsih, E. S., & Pratiwi, H. (2022). *Analisis Klaster K-Means Dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia*. *Seminar Nasional Official Statistics*, 2022(1), 1249–1258. <https://doi.org/10.34123/semnasoffstat.v2022i1.1511>