

## Optimisasi Produksi Budidaya Perikanan Di Provinsi Nusa Tenggara Timur Melalui Klasterisasi Menggunakan Algoritma *K-Means*

Annisa Maulana Sabilla<sup>1</sup>, Dinda Jelita<sup>2</sup>, Karimatul Aulia Hajijah<sup>3</sup>, Willdan Aprizal Arifin<sup>4</sup>

<sup>1,2,3,4</sup> Universitas Pendidikan Indonesia

Jl. Dr. Setiabudi, No.229, Kota Bandung, Jawa Barat 40154, Indonesia

e-mail: <sup>1</sup>annisamaulana@upi.edu, <sup>2</sup>dindaje19@upi.edu, <sup>3</sup>karimatul.aulia@upi.edu, <sup>4</sup>willdanarifin@upi.edu

Artikel Info : Diterima : 19-06-2024 | Direvisi : 13-11-2024 | Disetujui : 01-12-2024

**Abstrak** - Penurunan jumlah ikan di laut akibat *overfishing* dan perubahan kondisi iklim serta cuaca telah mengganggu aktivitas penangkapan ikan oleh nelayan. Hal ini menyebabkan ketimpangan antara hasil tangkapan laut dan produksi budidaya perikanan. Pada tahun 2018, hasil tangkapan laut di provinsi Nusa Tenggara Timur hanya mencapai 157.690,65 ton, jauh lebih rendah dibandingkan hasil budidaya yang mencapai 1.814.055,91 ton. Untuk mengoptimalkan produksi budidaya perikanan, penerapan pengelolaan yang tepat sangat diperlukan. Dalam penelitian ini, dilakukan studi literatur dan pengelompokan data menggunakan metode klasterisasi dengan algoritma *K-means*. Metode ini bertujuan untuk mengidentifikasi kelompok-kelompok wilayah atau jenis budidaya dengan karakteristik serupa sehingga pengelolaan sumber daya dapat lebih tepat sasaran. Penggunaan *Z-Score* dan pengukuran dengan tiga metode jarak, yaitu *Manhattan Distance*, *Minkowski Distance*, dan *Euclidean Distance*, dilakukan untuk menentukan klaster terbaik. Hasil penelitian menunjukkan bahwa nilai akurasi tertinggi berdasarkan *Silhouette Coefficient* (SC) adalah 0.1325489, yang diperoleh menggunakan *Minkowski Distance*. Hasil klasterisasi mengelompokkan satu wilayah masing-masing dalam cluster 1 dan 3, sementara tiga wilayah lainnya masuk dalam cluster 2.

Kata Kunci : *K-Means*, *Minkowski*, NTT

**Abstracts** - The paper's abstract templates are used as a guide for writing paper for publications in journals *Bina Sarana Informatika*. The paper writers must follow the instructions given in this paper. Abstract is restricted from 100-250 words in Indonesian Language and English Language (in italics), should not contain equations, pictures, and table. The letter size for abstract, keywords, and body papers is 10 points using Times New Roman font. In the abstract they contain prologue of the raised problem, the proposed method and the desired results (hypothesis). The number of pages in the text is a minimum of 5 and maximum of 10 pages using a paper A4 with a space of 1 (single).

Keywords : 3 words or phrases

### PENDAHULUAN

Nusa Tenggara Timur (NTT) dikenal sebagai salah satu provinsi dengan hasil tangkapan ikan terbesar di Indonesia. Dengan letaknya yang strategis di bagian timur kepulauan Nusa Tenggara, NTT dikelilingi oleh perairan yang luas dan kaya akan sumber daya ikan mencakup Laut Flores, Laut Sawu, dan Laut Timor. Kekayaan perairan Nusa Tenggara Timur ini tidak hanya mencakup hasil tangkapan ikan yang beragam jenisnya, tetapi juga budidaya perikanan yang substansial. Namun, keberagaman sumber daya tersebut tidak menjamin kelancaran aktivitas nelayan di wilayah tersebut. Para penyedia ikan atau nelayan di Nusa Tenggara Timur saat ini menghadapi berbagai kendala, terutama dalam sektor perikanan tangkap (Supriatin, et al., 2020). Jumlah ikan di laut yang terus menerus berkurang akibat *overfishing* dan kondisi iklim serta cuaca yang berubah-ubah, mengganggu aktivitas penangkapan ikan oleh nelayan (Azizi, Putri, & Fahrudin, 2017). Hal ini menciptakan ketimpangan antara hasil tangkapan dan produksi budidaya perikanan yang meliputi budidaya pembesaran, budidaya ikan hias, dan budidaya pembenihan. Sebagai contoh, pada tahun 2018, hasil tangkapan laut di provinsi Nusa Tenggara Timur hanya mencapai 157.690,65 ton, jauh lebih rendah dibandingkan dengan hasil budidaya pembesaran yang mencapai 1.814.055,91 ton.

Nusa Tenggara Timur memiliki peran yang sangat penting dalam budidaya perikanan nasional, terutama dalam budidaya ikan seperti tuna, tongkol, cakalang, dan jenis ikan lainnya. Selain tuna, tongkol, dan cakalang,



berbagai jenis ikan hias serta biota laut lainnya seperti rumput laut dan teripang juga dibudidayakan secara intensif di wilayah ini. Dalam pengelolaan suatu budidaya tentunya harus menggunakan teknik budidaya yang ramah lingkungan untuk meningkatkan efisiensi produksi sekaligus menjaga kelestarian lingkungan laut. Data yang dipublikasikan oleh Badan Pusat Statistik, menunjukkan bahwa dalam beberapa tahun terakhir, wilayah ini telah mencatat produksi budidaya perikanan dalam kategori alat tangkap di antaranya budidaya laut, tambak, kolam, keramba, jaring apung, dan sawah dengan jumlah total produksi pada tahun 2016 sebesar 1,841,934 ton, kemudian mengalami kenaikan pada tahun 2017 dengan hasil budidaya sebesar 1,953,524.54 ton, dan mengalami sedikit penurunan pada tahun 2018 menjadi 1,953,261 ton.

Dalam budidaya perikanan, pengelolaan yang efektif dan efisien penting untuk meningkatkan produksi dan kualitas ikan. Penurunan kualitas air dapat menyebabkan penurunan imunitas organisme sehingga rentan terhadap serangan penyakit (Jarir et al., 2020). Penggunaan *input* produksi yang tepat, pengelolaan faktor produksi, dan penerapan teknologi informasi dapat membantu mengoptimalkan hasil, meningkatkan produktivitas, dan menjaga kualitas air. Hal tersebut berlaku untuk tempat dan alat budidaya seperti budidaya laut, tambak, kolam, keramba, jaring apung, dan sawah, yang masing-masing memerlukan pengelolaan khusus sesuai kondisi spesifik. Misalnya, tambak dan kolam membutuhkan pengaturan kualitas air yang ketat, sementara jaring apung dan keramba memerlukan teknologi pemberian pakan otomatis. Dengan menerapkan pengelolaan yang tepat pada masing-masing metode budidaya ini, pembudidaya dapat memastikan bahwa setiap alat tangkap digunakan secara maksimal untuk menghasilkan produksi yang optimal dan berkualitas tinggi.

Klasterisasi merupakan salah satu metode dalam bidang *data mining* yang digunakan untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristiknya, dengan tujuan untuk mengidentifikasi kelompok-kelompok yang serupa dan membedakannya dari objek lain. Terdapat berbagai jenis analisis klasterisasi yang tersedia terdapat algoritma *k-means*, metode pengelompokan sekatan (*partitioning*), yang jelas lebih cepat daripada metode hierarkis dan lebih bermanfaat ketika jumlah objek sangat besar, adalah *K-means*. *K-means* adalah metode pengelompokan berbasis jarak yang membagi data menjadi beberapa klaster, dan algoritma ini hanya beroperasi pada atribut numerik. Algoritma *K-means* beroperasi dengan cara menentukan sejumlah *centroid*, di mana setiap data *point* akan ditempatkan pada klaster yang *centroid*-nya paling dekat dengannya. Proses ini dilakukan iteratif hingga posisi *centroid* tidak berubah atau perubahan yang terjadi sangat kecil.

Penggunaan metode *K-means* telah banyak digunakan dalam klasterisasi, seperti penelitian yang dilakukan oleh Ashari et al., (2022), tentang aplikasi data mining dengan metode *K-Means Clustering* dan *Index Davies Bouldin* untuk mengelompokkan film IMDB berdasarkan parameter seperti *runtime*, *IMDB rating*, *meta score*, jumlah suara, dan *gross*. Hasil penelitian menunjukkan bahwa rata-rata atribut tertinggi adalah 48.74 dan jumlah cluster yang terbentuk adalah 4 *cluster*. Lalu, perbandingan perhitungan jarak *Euclidean* dengan perhitungan jarak *Manhattan* pada *K-Means Clustering* di Kota Bekasi untuk membandingkan proses pengerjaan dan menentukan titik-titik pusat penyebaran penyakit *Covid* pada penelitian Cahya et al., (2023). Selain itu, pada penelitian yang dilakukan oleh Syakur et al., (2018) tentang Integrasi metode *K-Means Clustering* dan *Elbow Method* untuk identifikasi profil cluster pelanggan terbaik menggunakan kombinasi *K-Means* dengan *Elbow* untuk meningkatkan efisiensi dan efektivitas *K-Means* dalam mengelompokkan data pelanggan. Hasil penelitian menunjukkan bahwa metode ini dapat menghasilkan *cluster* yang lebih baik dan lebih akurat.

Pada budidaya perikanan, klasterisasi dapat membantu mengelompokkan wilayah produksi berdasarkan karakteristik tertentu, seperti jenis ikan yang dibudidayakan, metode budidaya yang digunakan, dan tingkat produksi. Dengan mengelompokkan data ini, dapat diidentifikasi pola-pola tertentu yang berguna dalam pengelolaan dan pengembangan sektor perikanan. Untuk mengoptimalkan produksi budidaya perikanan di Nusa Tenggara Timur, pengelompokan data menggunakan metode klasterisasi dapat menjadi langkah strategis. Klasterisasi dengan algoritma *K-means* akan membantu dalam mengidentifikasi kelompok-kelompok wilayah atau jenis budidaya yang memiliki karakteristik serupa sehingga pengelolaan sumber daya dapat lebih tepat sasaran. Dalam hal ini, penggunaan berbagai metode pengukuran jarak seperti *Euclidean*, *Manhattan*, dan *Minkowski* dalam algoritma *K-means* memungkinkan analisis yang lebih komprehensif.

## METODE PENELITIAN

### 1. Tahapan Metodologi Penelitian

Proses tahapan metodologi penelitian untuk *clustering* produksi budidaya perikanan:

#### a. Analisis Masalah

Mengidentifikasi permasalahan yang akan digunakan untuk pokok penelitian yaitu dilakukan pencarian pengukuran jarak terbaik memakai metode *euclidean distance*, *manhattan distance* dan *minkoswki distance* dengan penerapan *data minning k-means* pada pengelompokan dataset produksi budidaya ikan, sehingga dapat terbentuk suatu klaster yang mirip satu sama lain pada masing masing partisi.

#### b. Studi Literatur

Penelitian ini dibangun dengan dasar yang mengacu pada referensi-referensi yang mengulas informasi dan teori yang mendukungnya. Hal ini dilakukan dengan mengumpulkan materi dan informasi terkait algoritma klustering *K-means* dari berbagai jurnal dan penelitian terdahulu.

c. Pengumpulan Data

Pada tahap ini menentukan data yang akan diproses. Mencari data, memperoleh data yang dibutuhkan dan mengintegrasikan semua data ke dalam data set. Data yang akan digunakan adalah data produksi budidaya ikan Mengunjungi *website* resmi Badan Pusat Statistik Nusa Tenggara Timur untuk mendapatkan data laporan statistik produksi budidaya ikan. Pada penelitian ini memakai 3 jenis produksi perikanan budidaya menurut sub sektor dan kabupaten/ kota, yaitu Budidaya Laut, Kolam dan Tambak.

d. Pengolahan Data Awal

Data yang telah dikumpulkan selanjutnya di analisa dengan menyeleksi data yang meliputi pembersihan data dengan pembersihan data yang tidak lengkap (*missing value*). Mengelola data produksi perikanan budidaya provinsi Nusa Tenggara Timur menggunakan *MS. Excel* kemudian menyiapkan data tersebut untuk digunakan dalam penelitian.

e. Implementasi *K-Means*

Pada penelitian ini, akan dilakukan proses klustering dengan memakai algoritma *K-Means* pada data yang telah dianalisis dan akan menghasilkan bentuk *cluster* yang telah ditetapkan. Dan pada penelitian ini dilakukannya metode pengukuran jarak yaitu, *Euclidean distance*, *Manhattan distance* dan *Minkowski distance*.

f. Penarikan Kesimpulan

Proses menyimpulkan atau mengambil kesimpulan berdasarkan hasil analisis data yang telah dilakukan, dengan mempertimbangkan temuan-temuan signifikan yang ditemukan dalam penelitian.

**2. Analisa Data**

Data yang akan digunakan penulis untuk melakukan penelitian ini yaitu memakai data sekunder yang diperoleh dari Badan Pusat Statistik Nusa Tenggara Timur. Data set tersebut menampilkan *range* waktu dalam 3 tahun yaitu tahun 2016, 2017, 2018 dengan 23 wilayah kota/ kabupaten yang tertera. Pada dataset terdapat 6 sub sektor, dan pada penelitian ini kami memakai 3 sub sektor yaitu Budidaya Laut, Tambak dan kolam.

Tabel 1. Data Produksi Budidaya Perikanan Kabupaten/Kota

Wilayah	Produksi Perikanan Budidaya Menurut Sub sektor dan Kabupaten/Kota (Ton)								
	Budidaya Laut			Tambak			Kolam		
	2016	2017	2018	2016	2017	2018	2016	2017	2018
Sumba Barat	155.00	192720.00	192720.00	11.00	6.40	6.00	58.00	206.20	206.00
Sumba Timur	21546.00	123800.85	123801.00	-	-	-	11.00	108.78	109.00
Kupang	1342582.00	1343128.00	1343128.00	40.00	422.71	423.00	32.00	154.13	154.00
Timor Tengah Selatan	-	-	-	-	-	-	454.00	258.58	259.00
Timor Tengah Utara	-	-	-	151.00	1047.88	1048.00	11.00	230.39	230.00
Belu	-	-	-	134.00	116.60	117.00	34.00	26.20	26.00
Alor	161358.00	272.00	272.00	-	3.06	3.00	16.00	26.00	26.00
Lembata	7375.00	35.90	36.00	6.00	0.05	-	1.00	1.45	1.00
Flores Timur	124526.00	81466.19	81466.00	-	100.00	100.00	1.00	-	-
Sikka	20.00	-	-	-	-	-	1.00	0.05	-
Ende	26.00	1968.97	1969.00	13.00	29.66	30.00	23.00	31.47	31.00
Ngada	35.00	-	-	-	-	-	4.00	1396.71	1397.00
Manggarai	681.00	-	-	9.00	27.07	27.00	91.00	88.46	88.00
Rote Ndao	128595.00	126212.00	126212.00	3.00	1.65	2.00	9.00	3.84	4.00
Manggarai Barat	198.00	458.86	459.00	-	0.83	1.00	43.00	21.33	21.00
Sumba Tengah	10.00	273.47	273.00	32.00	-	-	4.00	9.27	10.00
Sumba Barat Daya	76.00	2401.60	2402.00	-	-	-	57.00	4.30	4.00
Nagekeo	160.00	14.98	15.00	-	48.50	48.00	4.00	10.21	10.00
Manggarai Timur	1312.00	74.93	75.00	20.00	3.97	4.00	92.00	142.33	142.00
Sabu Raijua	48214.00	68880.00	68880.00	-	-	-	-	-	-
Malaka	-	-	-	2665.00	6714.83	6715.00	865.00	173.96	174.00
Kota Kupang	-	-	-	2.00	7.73	-	52.00	-	8.00

Sumber : Hasil Penelitian (2024)

### 3. Data Mining

*Data mining* adalah proses analisis data yang bertujuan untuk mengungkap hubungan yang tidak terduga serta merangkum informasi dari data secara inovatif, sehingga memberikan pemahaman yang berharga dan bermanfaat bagi pemilik data. *Data mining* disebut sebagai proses menemukan pola dalam data, yang dapat dilakukan secara otomatis atau semi-otomatis, dengan syarat bahwa pola-pola yang ditemukan memiliki signifikansi atau relevansi yang penting.

*Data mining* melibatkan berbagai teknik dari statistik, pembelajaran mesin, dan kecerdasan buatan untuk mengeksplorasi dan menganalisis data dalam jumlah besar, dengan tujuan utama menemukan pola, hubungan, dan tren yang tersembunyi yang mungkin tidak langsung terlihat. Proses ini biasanya melibatkan beberapa tahap penting, seperti pengumpulan data, pembersihan data untuk memastikan kualitas yang lebih baik, transformasi data ke dalam format yang sesuai untuk analisis lebih lanjut, dan pemilihan sub set data yang relevan untuk analisis. Setelah itu, teknik *data mining* seperti klasifikasi, klustering, asosiasi, regresi, dan analisis deret waktu diterapkan untuk menemukan pola dalam data.

Setelah pola ditemukan, tahap evaluasi dan interpretasi dilakukan untuk memastikan pola tersebut valid, bermakna, dan dapat diterapkan dalam konteks praktis. Hasilnya kemudian disajikan dalam bentuk visual yang mudah dipahami, seperti grafik, diagram, dan peta, untuk membantu dalam interpretasi dan pengambilan keputusan.

### 4. Algoritma K-Means

Algoritma *K-Means* merupakan salah satu metode *clustering* yang paling terkenal dan banyak digunakan dalam berbagai aplikasi analisis data dan pembelajaran mesin. *Clustering* adalah teknik yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok (*clusters*) berdasarkan kemiripan antar data. Prinsip dasar dari *K-Means* adalah meminimalkan jarak antar data dalam *cluster* yang sama dan memaksimalkan jarak antar *cluster* yang berbeda.

Proses algoritma *K-Means* dimulai dengan menentukan *K* centroid awal, yang bisa dipilih secara acak atau menggunakan metode tertentu seperti *K-Means++* untuk pemilihan *centroid* yang lebih baik. Setelah inisialisasi, langkah pertama adalah langkah *assignment*, di mana setiap data di *assign* ke *centroid* terdekat berdasarkan jarak, biasanya jarak *Euclidean*. Selanjutnya, pada langkah *update*, *centroid* dihitung ulang sebagai rata-rata dari data yang telah di *assign* ke *centroid* tersebut. Proses ini, yang terdiri dari langkah *assignment* dan *update*, diulang secara *iteratif* hingga mencapai konvergensi, yaitu ketika tidak ada perubahan signifikan pada posisi *centroid*.

Algoritma *K-Means* terkenal karena kemudahan pemahaman dan implementasinya. Proses iteratif yang digunakan dalam *K-Means* juga relatif cepat dan efisien dalam banyak situasi. Selain itu, algoritma ini memiliki skalabilitas yang baik, mampu menangani dataset besar dengan kompleksitas komputasi yang rendah ( $O(nkt)$ , di mana  $n$  adalah jumlah data,  $k$  adalah jumlah *cluster*, dan  $t$  adalah jumlah iterasi). Namun, algoritma *K-Means* juga memiliki beberapa keterbatasan. Salah satu tantangan utamanya adalah penentuan jumlah *cluster*  $K$ ,  $K$  yang harus ditetapkan sebelumnya, sering kali tanpa informasi yang memadai. Selain itu, hasil akhir dari *K-Means* sangat bergantung pada inisialisasi *centroid*; inisialisasi yang kurang tepat dapat mengarah pada hasil *clustering* yang buruk. *K-Means* juga lebih efektif untuk *cluster* yang berbentuk bulat dan memiliki ukuran yang kurang lebih sama, sehingga kurang ideal untuk *cluster* dengan bentuk atau ukuran yang bervariasi. Selain itu, algoritma ini rentan terhadap *outlier*; data yang jauh dari kelompok utama dapat mempengaruhi hasil *clustering* secara signifikan.

### 5. Perhitungan Jarak

Langkah penting dalam pengelompokan adalah memilih metode perhitungan jarak yang tepat, karena ini dapat meningkatkan hasil dari algoritma yang berbasis jarak. Penting untuk mengidentifikasi bagaimana data tersebut saling berhubungan, seberapa mirip atau berbeda data yang ada satu sama lain, serta ukuran apa yang digunakan untuk membandingkannya. Tujuan utama perhitungan jarak adalah untuk memecahkan masalah tertentu dengan cara yang tepat. Dalam penelitian ini, metode perhitungan jarak yang digunakan adalah *Euclidean*, *Minkowski*, dan *Manhattan*.

#### a. Euclidean Distance

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$d$  adalah jarak antara titik data  $y$  dan pusat *cluster*  $x$ . Di mana  $x$  adalah pusat dari sebuah *cluster*,  $y$  adalah data pada atribut tertentu,  $i$  adalah setiap data dalam himpunan tersebut,  $n$  adalah jumlah total data,  $x_i$  adalah data yang merupakan pusat dari *cluster*  $i$ , dan  $y_i$  adalah data yang merupakan pusat dari data ke- $i$ .

#### b. Manhattan Distance

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|} \quad (2)$$

Rumus ini menghitung jarak antara dua titik, yaitu  $x$  dan  $y$ .  $d$  adalah jarak antara  $x$  dan  $y$ . Titik  $x$  merupakan pusat kluster, sedangkan  $y$  adalah data pada atribut tertentu. Indeks  $i$  menunjukkan setiap data, sementara  $n$  adalah jumlah total data. Nilai  $x_i$  adalah data pada pusat kluster  $i$  dan nilai  $y_i$  adalah data pada pusat data ke- $i$ .

c. *Minkowski Distance*

$$d(x, y) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (3)$$

Jarak *Minkowski* adalah generalisasi dari jarak *Euclidean* dan *Manhattan*. Dimana (X) dan (Y) adalah dua titik di dalam ruang,  $(x_i)$  dan  $(y_i)$  merupakan koordinat dari titik-titik tersebut,  $(n)$  adalah jumlah dimensi, dan  $(p)$  adalah orde dari metrik *Minkowski*. Ketika  $(p=2)$ , metrik ini menjadi jarak *Euclidean*, dan ketika  $(p=1)$ , metrik ini disebut jarak *Manhattan*.

## HASIL DAN PEMBAHASAN

Untuk analisis produksi perikanan, langkah-langkah yang diambil adalah sebagai berikut: Data produksi perikanan dikumpulkan dari berbagai sub sektor di berbagai wilayah. Lalu, data tersebut ditransformasi agar lebih mudah dianalisis. Hasil dari transformasi data ini kemudian ditampilkan dalam bentuk tabel untuk setiap sub sektor, yaitu Budidaya Laut, Tambak, dan Kolam di masing-masing wilayah. Tabel 2. menunjukkan hasil transformasi data produksi perikanan menurut sub sektor untuk berbagai wilayah.

Tabel 2. Hasil Transformasi Data

Wilayah	Budidaya Laut	Tambak	Kolam
Sumba Barat	385595.00	23.40	470.20
Sumba Timur	269147.85	0.00	228.78
Kupang	4028838.00	885.71	340.13
Timor Tengah Utara	0.00	2246.88	471.39
Belu	0.00	367.60	86.20
Alor	161902.00	6.06	68.00
Lembata	7446.90	6.05	3.45
Flores Timur	287458.19	200.00	1.00
Sikka	20.00	0.00	1.05
Ende	3963.97	72.66	85.47
Ngada	35.00	0.00	2797.71
Manggarai	681.00	63.07	267.46
Rote Ndao	381019.00	6.65	16.84
Manggarai Barat	1115.86	1.83	85.33
Sumba Tengah	556.47	32.00	23.27
Sumba Barat Daya	4879.60	0.00	65.30
Nagekeo	189.98	96.50	24.21
Manggarai Timur	1461.93	27.97	376.33
Sabu Raijua	185974.00	0.00	0.00
Malaka	0.00	16094.83	1212.96
Kota Kupang	0.00	9.73	60.00

Sumber : Hasil Penelitian (2024)

Setelah mendapatkan data transformasi pada analisis wilayah produksi perikanan menurut subsektor, lalu melakukan langkah selanjutnya yaitu melakukan proses normalisasi menggunakan metode *Z-Score* dilakukan untuk mengubah data menjadi bentuk yang memiliki rata rata nol dan standar deviasi satu. Hal ini bertujuan untuk mengurangi skala perbedaan antar variabel sehingga semua variabel memiliki bobot yang sama dalam analisis. Hasil normalisasi menunjukkan nilai yang sudah diubah berdasarkan rumus *Z-Score*, di mana nilai setiap data dikurangi dengan rata-rata dan dibagi dengan standar deviasi. Berikut merupakan hasil yang ditampilkan :



Tabel 3. Hasil Normalisasi Dengan Metode Z-Score

Wilayah	Budidaya Laut	Tambak	Kolam
Sumba Barat	-0.332905	-0.470550	1.083575
Kupang	1.777947	1.785202	0.420031
Ende	-0.554018	-0.341689	-0.879099
Rote Ndao	-0.335557	-0.514367	-1.229210
Manggarai Timur	-0.555467	-0.458596	0.604703

Sumber : Hasil Penelitian (2024)

Lalu selanjutnya langkah menentukan pusat *centroid* awal. Pusat *centroid* awal digunakan sebagai titik awal dalam algoritma *clustering*. Setiap wilayah (seperti Sumba Barat, Kupang, Ende, Rote Ndao, Manggarai Timur) memiliki jarak ke tiga *centroid* (C1, C2, C3) dihitung dengan tiga metode jarak yang berbeda: *Euclidean* (EUC), *Manhattan* (MAN), dan *Minkowski* (MIN).

Tabel 4. Pusat *Centroid* Awal

Cluster	Budidaya Laut	Tambak	Kolam
C1	1.777947	1.785202	0.420031
C2	-0.555467	-0.458596	0.604703
C3	-0.554018	-0.341689	-0.879099

Sumber : Hasil Penelitian (2024)

Setelah mendapatkan hasil pusat *centroid* awal lalu melakukan perhitungan jarak 3 metode. Langkah ini melibatkan perhitungan jarak antara titik data dan *centroid* menggunakan tiga metode jarak: *Euclidean*, *Manhattan*, dan *Minkowski*. Hasil dari perhitungan ini adalah koefisien *Silhouette* untuk setiap metode, yang digunakan untuk mengukur seberapa baik titik-titik data dikelompokkan. Metode dengan nilai koefisien *Silhouette* tertinggi dianggap memberikan hasil *clustering* terbaik.

Tabel 5. Hasil Perhitungan Jarak 3 Metode

Wilayah	Jarak Ke <i>Centroid</i>								
	EUC			MAN			MIN		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
Sumba Barat	3.159811e+00	5.281995e+01	9.79289e+00	5.030148e+00	0.713382	3.12648e+08	2.766596e+00	4.943906e+01	1.963794e+00
Kupang	5.858357e-07	3.242461e+00	3.413132e+00	9.225920e-07	4.761885	5.757987e+04	5.211995e-07	2.884820e+00	2.904208e+00
Ende	3.413131e+00	1.488401e+00	2.151344e-07	5.757986e+00	1.602153	5.509907e-08	2.904207e+00	1.484044e+00	1.885114e-07
Rote Ndao	3.531984e+00	1.847893e+00	4.473491e-01	6.062314e+00	2.109597	4.412512e-05	2.965793e+00	1.834984e+00	3.881769e-01
Manggarai Timur	3.242461e+00	7.541647e-07	1.488401e+00	4.761884e+00	0.000001	1.602158e+01	2.884820e+00	6.389442e-07	1.484044e+00

Sumber : Hasil Penelitian (2024)

Selanjutnya menentukan nilai akurasi dengan menggunakan metode SC, pada langkah ini melibatkan penentuan nilai akurasi dari metode *Silhouette Coefficient* untuk berbagai jumlah *cluster*. Nilai SC digunakan untuk mengevaluasi kekompakan dan pemisahan *cluster*. Pada tabel ini, akurasi dihitung dan dibandingkan untuk berbagai jumlah *cluster* untuk menentukan jumlah *cluster* yang optimal.

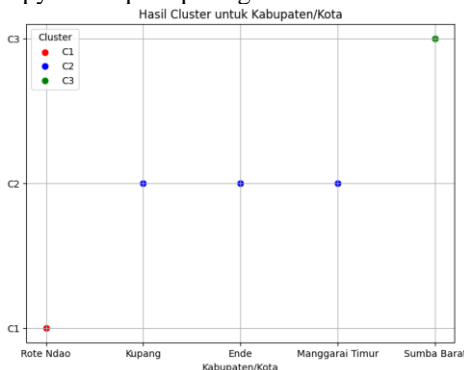
Tabel 6. Nilai Akurasi Metode SC

Akurasi	Jarak Ke <i>Centroid</i>			<i>Clusters</i>		
	C1	C2	C3	C1	C2	C3

Sumba Barat	2.766596e+00	4.943906e-01	1.963794e+00		*
Kupang	5.211995e-07	2.884820e+00	2.904208e+00		*
Ende	2.904207e+00	1.484044e+00	1.885114e-07		*
Rote Ndao	2.965793e+00	1.834984e+00	3.881769e-01	*	
Manggarai Timur	2.884820e+00	6.389442e-07	1.484044e+00		*

Sumber : Hasil Penelitian (2024)

Pada tahap selanjutnya, menentukan *cluster* yang sesuai dengan kategori wilayah berpotensi baik (BB) yang di simbolkan pada *cluster* 3, wilayah berpotensi sedang (BS) seperti pada *cluster* 2 dan wilayah kurang berpotensi seperti pada *cluster* 1. Adapun pada hasil pengelompokan sebelumnya ditampilkan dalam bentuk grafik *scatter* menggunakan bahasa pemrograman python seperti pada gambar di bawah ini.



Sumber : Hasil Penelitian (2024)

Gambar 1. Hasil Cluster

Pada grafik di atas menunjukkan bahwa wilayah Sumba Barat masuk dalam *cluster* 3 atau kategori wilayah berpotensi baik (BB), sedangkan Rote Ndao termasuk ke dalam *cluster* 1 yang berarti wilayah tersebut kurang berpotensi. Namun pada ketiga wilayah seperti Kupang, Ende, dan Manggarai Timur masuk ke dalam bagian *cluster* 2 yang di mana kategori tersebut termasuk dalam wilayah berpotensi sedang (BS).

## KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, didapat nilai akurasi menggunakan metode *silhouette coefficient* (SC) yaitu 0.1325489 sebagai jarak terbaik dari hasil pengukuran pada metode *Minkowski Distance*. Metode tersebut kemudian digunakan sebagai acuan untuk pengelompokan hasil *cluster* yang diperoleh di mana terdapat masing-masing 1 wilayah pada golongan *cluster* 1 dan 3, sedangkan 3 wilayah lainnya tergolong dalam *cluster* 2. Hal ini berarti wilayah berpotensi sedang yang mendominasi di Provinsi Nusa Tenggara Timur dan untuk wilayah yang tergolong ke dalam kategori kurang berpotensi sangat memerlukan perhatian dan langkah yang serius dari pemerintah agar dapat mengoptimalkan produksi budidaya perikanan yang dilakukan. Penelitian ini menunjukkan bahwa penerapan algoritma *K-means* untuk klasterisasi produksi budidaya perikanan di Provinsi Nusa Tenggara Timur (NTT) sangat efektif dalam mengidentifikasi kelompok-kelompok wilayah atau jenis budidaya dengan karakteristik serupa. Dengan menggunakan metode pengukuran jarak *Euclidean*, *Manhattan*, dan *Minkowski*, proses klasterisasi dapat dilakukan secara komprehensif. Hasil penelitian ini dapat membantu para pembudidaya dan pengelola perikanan untuk mengoptimalkan pengelolaan sumber daya perikanan mereka, sehingga produksi dapat lebih maksimal dan berkualitas tinggi. Klasterisasi ini juga membantu dalam merencanakan strategi pengembangan budidaya perikanan yang lebih tepat sasaran berdasarkan karakteristik masing-masing wilayah.

## REFERENSI

- Azizi, A., Putri, E. I. K., & Fahrudin, A. (2017). Analisis faktor-faktor yang mempengaruhi perubahan pendapatan nelayan akibat variabilitas iklim. *Jurnal Sosial Ekonomi Kelautan Dan Perikanan*, 12(2), 225-233.
- Arthur, D., & Vassilvitskii, S. (2007). K-means++: The Advantages of Careful Seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*.
- Ashari, I.F., Banjarnahor, R., Farida, D.R., Aisyah, S.P., Dewi, A.P., & Humaya, N. (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing*.

- Br.Giniting, N. S., Maulita, Y., & Ramadan, S. (2023). Pengelompokan pasien operasi sesar berdasarkan indikasinya menggunakan metode clustering. *Journal of Engineering, Technology and Computing (JETCom)*, 2(3).
- Cahya, F.N., Mahatma, Y., & Rohimah, S.R. (2023). Perbandingan Metode Perhitungan Jarak Euclidean dengan Perhitungan Jarak Manhattan pada K-Means Clustering Dalam Menentukan Penyebaran Covid di Kota Bekasi. *JMT : Jurnal Matematika dan Terapan*.
- Fajriah, R. I., Sutisna, H., & Simpony, B. K. (2019). Perbandingan Distance Space Manhattan Dengan Euclidean Pada K-Means Clustering Dalam Menentukan Promosi. *JCIT (Indonesian Journal on Computer and Information Technology)*, 4(1), 36-49.
- Jarir, D. V., Anton, A., Anton, S. W., Yunarti, Y., Fatmah, F., Jayadi, J., & Usman, H. (2020). Strategi pengelolaan tambak udang vaname (*Litopenaeus vannamei*) terhadap sebaran penyakit parasiter di Kecamatan Tanete Riattang Timur. *JOURNAL OF INDONESIAN TROPICAL FISHERIES (JOINT-FISH): Jurnal Akuakultur, Teknologi dan Manajemen Perikanan Tangkap dan Ilmu Kelautan*, 3(1), 28-39.
- Kaamilah, L. L., & Mulyawan. (2023). Analisis Kelompok Lansia Berdasarkan Kategori Usia Dengan Metode K-Means Clustering. *Jurnal Riset Ilmu Akuntansi*, 2(2), 01-14.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 4(1), 20.
- Syakur, M.A., Khotimah, B.K., Rochman, E.M., & Satoto, B.D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336.
- Supriatin, F. E., Rohman, A. N., Akreditasi, B., Perguruan, N., Kantor, T., Pendidikan, K., & Kebudayaan, D. (2020). Peramalan Produksi Perikanan Budidaya di Kabupaten Malang Dengan Metode Exponential Smoothing. *Media Akuatika J. Ilm. Jur. Budid. Perair*, 5(2), 51-58.
- Widodo, S., Brawijaya, H., & Samudi. (2021). Clustering Kanker Serviks Berdasarkan Perbandingan Euclidean dan Manhattan Menggunakan Metode K-Means. *Jurnal Media Informatika Budidarma*, 5(2), 687-694.