

Implementasi Algoritma *K-Means* untuk Klusterisasi Data Hasil Tangkapan Ikan di Karimunjawa

¹Muchammad Dwi Rivaldo, ²Gentur Wahyu Nyipto Wibowo, ³Harminto Mulyo

¹Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Nahdlatul Ulama Jepara, Indonesia.

portfoliorivaldo@gmail.com, gentur@unisnu.ac.id, minto@unisnu.ac.id

ABSTRAK

Penelitian ini berfokus pada implementasi algoritma K-Means untuk klusterisasi data hasil tangkapan ikan di Karimunjawa. Tujuan utama adalah untuk mengidentifikasi jenis ikan yang paling banyak ditangkap dan area penangkapan yang paling produktif. Data yang digunakan mencakup tanggal, nama nelayan, jumlah ikan (kg), jenis ikan, area penangkapan ikan, dan metode penangkapan ikan, yang dikumpulkan selama tahun 2020. Analisis klusterisasi menghasilkan tiga klaster utama: Klaster 0 dengan total 315,9 kg terdiri dari Cumi, Kakak Tua, Jinahak, Baronang, Panti, dan Tambak Jeron; Klaster 1 dengan total 856,9 kg terdiri dari Teri, Tengiri, Tambak Jeron, Udul, Panti, dan Pari; Klaster 2 dengan total 1383,2 kg terdiri dari Todak. Selain itu, area penangkapan yang produktif juga diklusterisasi menjadi tiga: Klaster 0 mencakup Karimunjawa Timur, Klaster 1 mencakup Karimunjawa Barat, dan Klaster 2 mencakup Karimunjawa Utara. Hasil evaluasi menggunakan metrik pengukuran menunjukkan bahwa *Silhouette Score* positif sebesar 0,48 mengindikasikan bahwa klaster yang dihasilkan cenderung terpisah dengan baik, meskipun masih ada ruang untuk perbaikan. *Davies-Bouldin Index* yang rendah sebesar 0,83 menunjukkan bahwa klaster yang dihasilkan cukup terpisah satu sama lain, meskipun tidak sempurna. *Metode Elbow* memberikan indikasi jumlah klaster optimal, membantu dalam pemilihan konfigurasi yang tepat untuk analisis klaster. Penelitian ini memberikan wawasan berharga tentang distribusi tangkapan ikan di Karimunjawa, yang dapat digunakan untuk meningkatkan strategi penangkapan dan manajemen perikanan. Implementasi dan analisis dilakukan menggunakan Google Colab dengan bahasa pemrograman Python.

Kata Kunci: algoritma *k-means*; data mining; nelayan Karimunjawa; perikanan

ABSTRACT

This research focuses on the implementation of the K-Means algorithm for clustering fish catch data in Karimunjawa. The primary objective is to identify the most commonly caught fish species and the most productive fishing areas. The data used includes the date, fisherman's name, fish quantity (kg), fish species, fishing area, and fishing method, collected throughout the year 2020. The clustering analysis yielded three main clusters: Cluster 0 with a total of 315.9 kg consisting of squid, parrotfish, snapper, rabbitfish, panti, and tambak jeron; Cluster 1 with a total of 856.9 kg consisting of anchovy, mackerel, tambak jeron, udul, panti, and ray; Cluster 2 with a total of 1383.2 kg consisting of swordfish. Additionally, the productive fishing areas were also clustered into three: Cluster 0 covers East Karimunjawa, Cluster 1 covers West Karimunjawa, and Cluster 2 covers North Karimunjawa. The evaluation results using measurement metrics show a positive *Silhouette Score* of 0.48, indicating that the clusters tend to be well-separated, although there is still room for improvement. A low *Davies-Bouldin Index* of 0.83 indicates that the clusters

are fairly well separated from each other, though not perfectly. The Elbow method provided an indication of the optimal number of clusters, assisting in the selection of the appropriate configuration for cluster analysis. This research provides valuable insights into the distribution of fish catches in Karimunjawa, which can be used to improve fishing strategies and fisheries management. The implementation and analysis were carried out using Google Colab with the Python programming language.

Keywords: *K-Means algorithm; data mining; Karimunjawa fishermen; fisheries*

PENDAHULUAN

Perikanan memainkan peran penting dalam perekonomian dan ketahanan pangan global, khususnya di negara kepulauan seperti Indonesia. Karimunjawa, dengan keanekaragaman hayati lautnya yang kaya, merupakan salah satu daerah yang memiliki potensi besar dalam sektor perikanan. Namun, wilayah ini menghadapi berbagai tantangan dalam pengelolaan sumber daya perikanan, seperti overfishing, degradasi habitat, keterpencilan geografis, sarana transportasi yang terbatas, Illegal, Unreported and Unregulated Fishing (IUUF), keterbatasan jumlah dan ukuran alat tangkap, serta kurangnya kerjasama terpadu antar pemangku kepentingan (Wardono et al. 2019). Salah satu permasalahan spesifik adalah kesulitan dalam mengidentifikasi jenis ikan yang banyak ditangkap dan area penangkapan ikan yang produktif berdasarkan data historis tangkapan.

Dalam upaya meningkatkan pengelolaan perikanan yang berkelanjutan, analisis data hasil tangkapan ikan dapat memberikan wawasan yang berharga. Algoritma *K-Means* telah terbukti efektif dalam mengelompokkan data hasil tangkapan ikan, seperti yang ditunjukkan dalam penelitian sebelumnya yang mengelompokkan hasil tangkapan ikan selama periode 2015-2017 menjadi dua kluster: satu untuk hasil tangkapan rendah dan satu lagi untuk hasil tangkapan tinggi. Penelitian tersebut menemukan bahwa ikan layang dan cakalang merupakan jenis ikan unggulan yang dikategorikan dalam kluster dengan hasil tangkapan tinggi (Julianti Hablum et al. 2019).

Pendekatan klusterisasi juga telah diterapkan dalam berbagai konteks lainnya, seperti pengelompokan data kasus Covid-19 di Provinsi Lampung yang menunjukkan variasi pengelompokan berdasarkan atribut kesehatan dan demografi (Nabila, Rahman Isnain, and Abidin 2021). Selain itu, analisis klusterisasi digunakan untuk menilai kinerja dosen berdasarkan frekuensi pertemuan mata kuliah, yang memberikan wawasan berharga bagi pengambilan kebijakan dalam lembaga pendidikan (Virgo, Defit, and Yuhandri 2020). Penelitian oleh Luan (2020) membahas implementasi optimasi algoritma *K-Means* pada sistem yang dapat dikonfigurasi ulang. Studi ini memberikan solusi akselerasi hardware yang efektif untuk algoritma pembelajaran mesin pada data besar, meningkatkan kecepatan dan efisiensi pemrosesan data (Luan 2020). Pada penelitian yang dilakukan oleh Xie et al. (2020) memfokuskan penelitian mereka pada peningkatan algoritma *K-Means* untuk klusterisasi data besar dengan mempertahankan akurasi klusterisasi dan meningkatkan efisiensi melalui klusterisasi ruang fitur dan proses *K-Means* yang dipercepat. Hasil eksperimen menunjukkan bahwa pendekatan ini secara signifikan meningkatkan performa algoritma pada skenario data besar (Xie, Liu, and Wei 2020). Menurut penelitian Lacetti et al., (2020) ada dua masalah utama algoritma *K-Means*: kesulitan dalam menentukan nilai *K* dan eksekusi algoritma pada arsitektur tradisional. Kelemahan ini sering menyebabkan hasil clustering yang tidak optimal pada masalah nyata (Lacetti et al. 2020).

Penelitian ini bertujuan untuk menerapkan algoritma *K-Means* dalam klusterisasi data hasil tangkapan ikan di Karimunjawa. Dengan mengidentifikasi pola dan tren dalam data hasil tangkapan, penelitian ini diharapkan dapat memberikan wawasan yang berguna bagi nelayan dan pengelola perikanan dalam mengoptimalkan strategi penangkapan serta mendukung upaya konservasi dan pengelolaan sumber daya laut yang berkelanjutan. Penelitian ini juga bertujuan untuk menunjukkan potensi dan manfaat penggunaan metode klusterisasi dalam analisis data

perikanan.

TINJAUAN PUSTAKA

1. Data Mining

Data mining adalah proses penemuan pola, hubungan, atau informasi berguna dari sejumlah besar data yang sering kali tersimpan dalam basis data, data warehouse, atau bentuk penyimpanan data lainnya. Proses ini melibatkan berbagai teknik dari bidang statistik, pembelajaran mesin (machine learning), kecerdasan buatan (artificial intelligence), dan sistem basis data untuk mengekstraksi pengetahuan dari data yang besar dan kompleks.

Menurut Putri Ariyani et al. (2023) proses ekstraksi pola yang bermakna dari volume data yang sangat besar dikenal sebagai data mining. Pola yang diidentifikasi harus memiliki signifikansi dan memberikan manfaat, yang sering kali berupa keuntungan ekonomi. Data yang digunakan dalam proses ini dapat disimpan dalam basis data, data warehouse, atau berbagai bentuk penyimpanan data lainnya. (Apriyani, Dikananda, and Ali 2023)

Nabila et al. (2021) menekankan pentingnya penemuan pengetahuan atau knowledge discovery dalam data mining, di mana tujuan utama adalah mengungkap informasi tersembunyi dalam data besar (Nabila et al. 2021)

2. Algoritma *K-Means*

Penelitian sebelumnya Prasetyo et al., (2023) menjelaskan bahwa algoritma adalah serangkaian langkah atau prosedur yang jelas dan terdefinisi dengan baik untuk menyelesaikan suatu masalah atau mencapai tujuan tertentu. Algoritma biasanya digunakan dalam konteks pemrograman dan ilmu komputer, tetapi konsep ini juga berlaku dalam berbagai disiplin ilmu lainnya. *K-Means* Clustering adalah salah satu algoritma Machine Learning yang populer dalam kategori Unsupervised Learning. Algoritma ini mengharuskan pengguna untuk menentukan nilai *k*, yaitu jumlah klaster yang akan dibentuk, sebelum proses klasterisasi dimulai. Nilai *k* tersebut merujuk pada jumlah centroid atau titik pusat dari masing-masing klaster (Prasetyo, Pranoto, and Primaswara Prasetya 2023).

Menurut Elda et al., (2021) menekankan bahwa klasterisasi menggunakan algoritma *K-Means* adalah salah satu metode yang digunakan dalam Data Mining. Klaster adalah sekumpulan objek data yang memiliki kesamaan satu sama lain dalam kelompok yang sama, namun berbeda dari objek di kelompok lainnya. Klasterisasi termasuk dalam kategori unsupervised learning yang bertujuan untuk mempartisi data yang tidak memiliki label ke dalam beberapa kelompok (Elda et al. 2021)

Dalam penelitian Hayati et al., (2023) berhasil menerapkan metode data mining *K-Means* clustering untuk menganalisis pola ekspor Indonesia selama periode 2020 hingga 2022. Hasil analisis menunjukkan adanya tiga klaster yang mencerminkan tingkat ekspor barang, yaitu klaster rendah (C0) dengan 99 data, klaster sedang (C1) dengan 3 data, dan klaster tinggi (C2) dengan 1 data. Dengan nilai Davies-Bouldin Index (DBI) terkecil sebesar 0.3665 untuk tiga klaster, penelitian ini memberikan wawasan mendalam mengenai pola ekspor (Hayati 2023).

Berikut rumus dari algoritma *K-Means*:

$$d(x, y) = \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} \dots \dots \dots 2$$

Keterangan:

d = titik dokumen

x = data record

y = data centroid

Jarak terpendek antara centroid dan dokumen menentukan penempatan dokumen tersebut dalam sebuah cluster. Misalnya, jika dokumen A memiliki jarak terpendek ke centroid 1 dibandingkan dengan centroid lainnya, maka dokumen A akan dimasukkan ke dalam cluster 1. Selanjutnya, posisi centroid baru untuk masing-masing centroid (*C_{i-j}*) dihitung kembali dengan mengambil rata-rata dokumen yang termasuk dalam cluster awal (*G_{i-j}*). Proses iterasi ini dilakukan secara berulang hingga posisi cluster tidak mengalami perubahan (Julianti Hablum et al. 2019)

Berikut rumus dari penentuan centroid:

$$C(i) = \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{\sum x} \dots\dots\dots 3$$

Keterangan:

x_1 = nilai data *record* ke-1

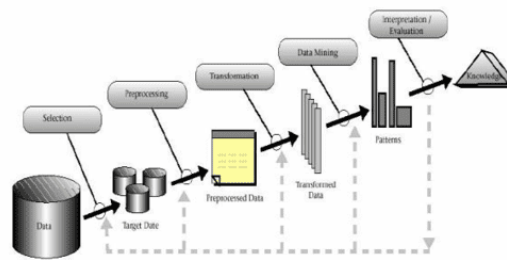
x_2 = nilai data *record* ke-2

$\sum x$ = jumlah data *record*

METODE PENELITIAN

3.1. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah Knowledge Discovery in Database (KDD), yang bertujuan untuk mengekstraksi pengetahuan dari kumpulan data besar. Ketika menerapkan teknik data mining ini, proses analisis data dilakukan melalui tahapan-tahapan yang ditentukan oleh KDD (Ejimofor and Okonkwo 2021)



Gambar 1: menggambarkan proses Knowledge Discovery in Databases (KDD) yang terdiri dari beberapa tahapan sebagai berikut:

1. Pemilihan Data (Data Selection)

Pada tahap awal ini, data yang relevan dan diperlukan dipilih untuk memulai proses KDD. Data yang telah dipilih ini akan digunakan pada tahap-tahap berikutnya dan disimpan secara terpisah dari database operasional (Pilnenskiy and Smetannikov 2020).

2. Praproses atau Pembersihan (Preprocessing or Cleaning)

Sebelum memulai data mining, langkah awal yang harus dilakukan adalah membersihkan data. Tujuan dari tahap ini adalah untuk menghapus data duplikat, memeriksa konsistensi data, dan memperbaiki kesalahan seperti kesalahan pengetikan atau cetak. Proses enrichment juga dilakukan untuk menambahkan informasi tambahan yang relevan ke dalam data (Jo et al. 2022).

3. Transformasi (Transformation)

Pada tahap ini, data yang telah dipilih diubah agar sesuai untuk digunakan dalam proses data mining. Langkah ini sangat bergantung pada jenis informasi yang ingin diperoleh dari database.

4. Data Mining

Tahap ini merupakan inti dari proses, di mana pola atau informasi menarik diidentifikasi dari data yang telah diproses menggunakan teknik atau metode tertentu. Pemilihan metode dan algoritma yang tepat sangat penting agar sesuai dengan tujuan keseluruhan dari proses KDD (Apriyani et al. 2023).

5. Interpretasi dan Evaluasi (Interpretation/Evaluation)

Pola atau informasi yang diperoleh dari data mining harus disampaikan dengan jelas kepada pemangku kepentingan agar mudah dipahami. Langkah ini juga mencakup evaluasi untuk memastikan bahwa pola yang ditemukan sesuai atau tidak dengan fakta dan hipotesis yang ada

sebelumnya. Dengan mengikuti tahapan-tahapan ini, proses KDD bertujuan untuk menemukan pengetahuan yang berharga dan tersembunyi dalam kumpulan data besar, sehingga dapat mendukung pengambilan keputusan yang lebih baik.

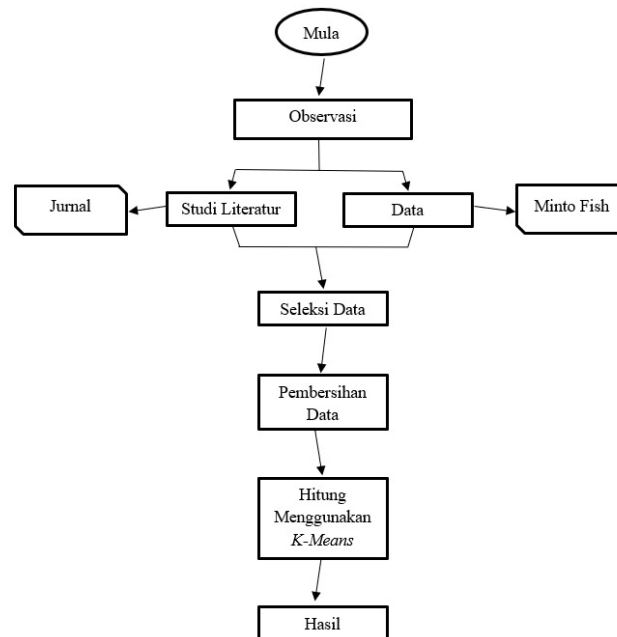


Diagram 1: Tahapan Penelitian

3.2. Sumber Data

Dalam penelitian ini, sumber data yang digunakan berasal dari nelayan desa Karimunjawa dan pengepul ikan yang ada di desa Karimunjawa, Jepara, Jawa Tengah, Indonesia. Data yang dikumpulkan meliputi hasil tangkapan ikan, jenis ikan, lokasi penangkapan, waktu penangkapan, serta teknik penangkapan ikan. Kepercayaan pada data yang diperoleh dari para nelayan dan pengepul ikan ini menjadi kunci penting dalam penelitian ini, karena data ini merupakan referensi utama untuk analisis clustering yang akan dilakukan. Dengan memanfaatkan data lapangan ini, penelitian ini diharapkan dapat memberikan wawasan yang berharga tentang karakteristik hasil tangkapan ikan di Karimunjawa, yang dapat menjadi dasar untuk pengambilan keputusan yang lebih baik dalam pengelolaan dan pengembangan sektor perikanan di wilayah tersebut. Data yang diambil mencakup periode penangkapan ikan pada tahun 2023.

3.3. Populasi dan Sampel

Populasi dalam penelitian ini didefinisikan sebagai sekumpulan objek atau subjek yang memiliki karakteristik tertentu, yang menjadi dasar bagi peneliti dalam menarik kesimpulan. Pada tahun 2023, kelompok utama yang menjadi fokus dalam bidang perikanan adalah populasi yang dibahas dalam studi ini. Proses sampling merupakan metode untuk mengambil sebagian kecil dari populasi, sehingga dapat mewakili nilai atau karakteristik keseluruhan. Bagian yang diambil dari populasi ini dikenal sebagai sampel, yang berfungsi sebagai representasi dari keseluruhan populasi. Dalam konteks penelitian ini, populasi mencakup seluruh hasil tangkapan ikan di Desa Karimunjawa, Kabupaten Jepara, Provinsi Jawa Tengah.

Sampel yang digunakan dalam penelitian ini merupakan bagian dari populasi yang lebih besar. Mengingat bahwa menganalisis seluruh data hasil tangkapan dari semua nelayan di Desa Karimunjawa akan menjadi tugas yang sangat kompleks, sampel ini akan dimanfaatkan untuk melakukan analisis clustering menggunakan algoritma *K-Means*. Dengan memilih sampel yang representatif, penelitian ini bertujuan untuk memberikan wawasan yang bermakna mengenai pola dan tren dalam hasil tangkapan ikan di Desa Karimunjawa, tanpa perlu menganalisis seluruh

populasi yang sangat luas.

3.4. Teknik Pengumpulan Data

Metode penelitian kepustakaan yang diterapkan dalam studi ini melibatkan proses analisis, pencarian, dan pengumpulan informasi yang relevan dengan tema penelitian. Proses ini berfungsi sebagai dasar dalam pengumpulan data untuk penelitian ini. Data yang digunakan untuk mengelompokkan hasil produksi perikanan budidaya berdasarkan jenis ikan diperoleh dari nelayan dan pengepul ikan di Desa Karimunjawa pada tahun 2023. Selanjutnya, untuk mengklasifikasikan produksi perikanan budidaya sesuai dengan jenis ikan unggulan, data yang telah dikumpulkan akan dianalisis menggunakan metode *K-Means*, dengan mempertimbangkan beberapa atribut yang terdapat dalam dataset tersebut.

3.5. Teknik Analisis Data

Analisis data merupakan proses yang sistematis dan terstruktur yang bertujuan untuk mengumpulkan serta mengorganisir informasi dari berbagai sumber. Langkah-langkah dalam analisis ini mencakup pengelompokan data ke dalam kategori yang relevan, penyajian data dalam unit-unit yang sesuai, dan penyusunan informasi secara menyeluruh. Selain itu, proses ini bertujuan untuk mengidentifikasi pola-pola yang muncul, menyoroti informasi yang signifikan, menentukan fokus penelitian, dan merangkum hasil analisis untuk mendapatkan pemahaman yang lebih mendalam mengenai fenomena yang diteliti serta konteks sosial di sekitarnya. Dengan pendekatan ini, analisis data dapat memberikan wawasan yang berharga untuk pengambilan keputusan yang lebih tepat dan informatif.

HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Data Selection

Data selection adalah proses penting dalam menentukan data mana yang akan digunakan untuk analisis, serta data mana yang tidak relevan. Dalam konteks penggunaan Google Colab dengan bahasa pemrograman Python, langkah pertama dalam persiapan analisis adalah memuat dataset yang diperlukan. Setelah data dimuat, pengguna dapat menggunakan berbagai teknik untuk memilih subset data yang relevan, seperti filtering berdasarkan kolom tertentu atau kondisi spesifik. Berikut langkah data selection:

1. Import Data ke Google Colab ('data_tangkapan_ikan.xlsx')
2. Menampilkan beberapa data baris pertama untuk memastikan data sudah benar terbaca oleh program.

```
import pandas as pd

# Load dataset
data = pd.read_excel('data_tangkapan_ikan.xlsx')

# Display the first few rows of the dataset
print(data.head())
```

	Tanggal	Nama Nelayan	Jumlah Ikan (Kg)	Jenis Ikan	\
0	2023-02-03 00:00:00	Siswanto	2.8	Cumi	
1	2023-02-03 00:00:00	Misro	2.1	Cumi	
2	2023-02-04 00:00:00	Saiful	10.0	Teri	
3	2023-02-03 00:00:00	Man	10.0	Tengiri	
4	2023-02-04 00:00:00	Man	0.6	Cumi	

	Area Penangkapan	Metode Penangkapan
0	Karimunjawa Timur	Pancing
1	Karimunjawa Timur	Pancing
2	Karimunjawa Barat	Jaring
3	Karimunjawa Utara	Jaring
4	Karimunjawa Timur	Pancing

Gambar 2: Import Data dan Menampilkan Data

4.1.2. Preprocessing/ Data Cleaning

Data cleaning adalah tahap krusial dalam proses analisis data yang bertujuan untuk memastikan bahwa dataset bebas dari kesalahan, duplikasi, dan ketidakakuratan. Dalam hal ini peneliti melakukan:

1. Menghapus data yang kosong
2. Mengkodekan variabel kategori ('Jenis Ikan', 'Area Penangkapan', dan 'Metode Penangkapan') menjadi numerik.

```
# Handling missing values (if any)
data_cleaned = data.dropna()

# Encoding categorical variables
data_cleaned['Jenis Ikan'] = data_cleaned['Jenis Ikan'].astype('category').cat.codes
data_cleaned['Area Penangkapan'] = data_cleaned['Area Penangkapan'].astype('category').cat.codes
data_cleaned['Metode Penangkapan'] = data_cleaned['Metode Penangkapan'].astype('category').cat.codes

# Display cleaned data
print(data_cleaned.head())
```

	Tanggal	Nama Nelayan	Jumlah Ikan (Kg)	Jenis Ikan \
0	2023-02-03 00:00:00	Siswanto	2.8	1
1	2023-02-03 00:00:00	Misro	2.1	1
2	2023-02-04 00:00:00	Saiful	10.0	9
3	2023-02-03 00:00:00	Man	10.0	8
4	2023-02-04 00:00:00	Man	0.6	1

	Area Penangkapan	Metode Penangkapan
0	4	1
1	5	1
2	0	0
3	7	0
4	4	1

Gambar 3: Preprocessing Data/ Data Cleaning

4.1.3. Transformasi

Transformasi data adalah tahap penting dalam proses analisis yang bertujuan untuk memodifikasi dan menyiapkan dataset agar sesuai dengan kebutuhan analisis. Dalam hal ini peneliti melakukan:

1. Memilih fitur yang akan digunakan untuk clustering ('Jumlah Ikan (Kg)', 'Jenis Ikan', 'Area Penangkapan').
2. Menstandarisasi fitur-fitur tersebut.

```
from sklearn.preprocessing import StandardScaler

# Select features for clustering
features = data_cleaned[['Jumlah Ikan (Kg)', 'Jenis Ikan', 'Area Penangkapan']]

# Standardize features
scaler = StandardScaler()
features_scaled = scaler.fit_transform(features)

# Display scaled features
print(features_scaled[:5])
```

```
[[-0.2632064 -0.98863091 0.27851238]
 [-0.27815613 -0.98863091 0.83885278]
 [-0.10943765 1.00040033 -1.96284918]
 [-0.10943765 0.75177142 1.95953356]
 [-0.31019129 -0.98863091 0.27851238]]
```

Gambar 4: Transformasi Data

4.1.4. Data Mining

Data mining adalah proses yang bertujuan untuk menemukan pola dan informasi berharga

dari dataset yang besar dan kompleks. Penelitian ini menggunakan algoritma *K-Means* clustering dengan bahasa pemrograman python. Dalam hal ini peneliti melakukan:

1. Menggunakan algoritma *K-Means* dengan 3 kluster.
2. Menambahkan kolom kluster ke data yang sudah dibersihkan.

```

from sklearn.cluster import KMeans

# Apply K-Means clustering
kmeans = KMeans(n_clusters=3, random_state=0)
data_cleaned['Cluster'] = kmeans.fit_predict(features_scaled)

# Display clustered data
print(data_cleaned.head())

```

	Tanggal	Nama Nelayan	Jumlah Ikan (Kg)	Jenis Ikan
0	2023-02-03 00:00:00	Siswanto	2.8	1
1	2023-02-03 00:00:00	Misro	2.1	1
2	2023-02-04 00:00:00	Saiful	10.0	9
3	2023-02-03 00:00:00	Man	10.0	8
4	2023-02-04 00:00:00	Man	0.6	1

	Area Penangkapan	Metode Penangkapan	Cluster
0	4	1	0
1	5	1	0
2	0	0	1
3	7	0	1
4	4	1	0

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_warnings.warn`

Gambar 5: Data Mining menggunakan Algoritma *K-Means* Clustering

4.1.5. Evaluasi

Evaluasi hasil clustering adalah tahap penting dalam proses data mining yang bertujuan untuk menilai kualitas dan keandalan hasil pengelompokan yang telah dilakukan. Dalam studi ini, scatter plot digunakan untuk mengelompokkan data tangkapan ikan berdasarkan kluster yang telah dihasilkan menggunakan algoritma *K-Means* clustering. Dengan menggunakan scatter plot, data yang telah di-scaling ditampilkan dalam dua dimensi, dimana sumbu x mewakili jumlah ikan yang ditangkap (kg), dan sumbu y mewakili jenis ikan. Setiap titik pada plot ini mewakili satu entri data tangkapan ikan, dan warnanya menunjukkan kluster mana titik tersebut termasuk. Warna yang berbeda digunakan untuk menunjukkan berbagai kluster, memberikan visualisasi yang jelas tentang bagaimana data dikelompokkan. Peneliti melakukan:

1. Visualisasi hasil clustering dalam bentuk scatter plot.
2. Mengelompokkan data berdasarkan kluster dan menampilkan informasi tentang total jumlah ikan, jenis-jenis ikan, dan area penangkapan untuk masing-masing kluster.

```

# Visualisasi hasil clustering
plt.figure(figsize=(10, 6))
plt.scatter(features_scaled[:, 0], features_scaled[:, 1], c=data_cleaned['Cluster'], cmap='viridis')
plt.xlabel('Jumlah Ikan (Kg)')
plt.ylabel('Jenis Ikan')
plt.title('K-Means Clustering of Fish Catch Data')
plt.colorbar(label='Cluster')
plt.show()

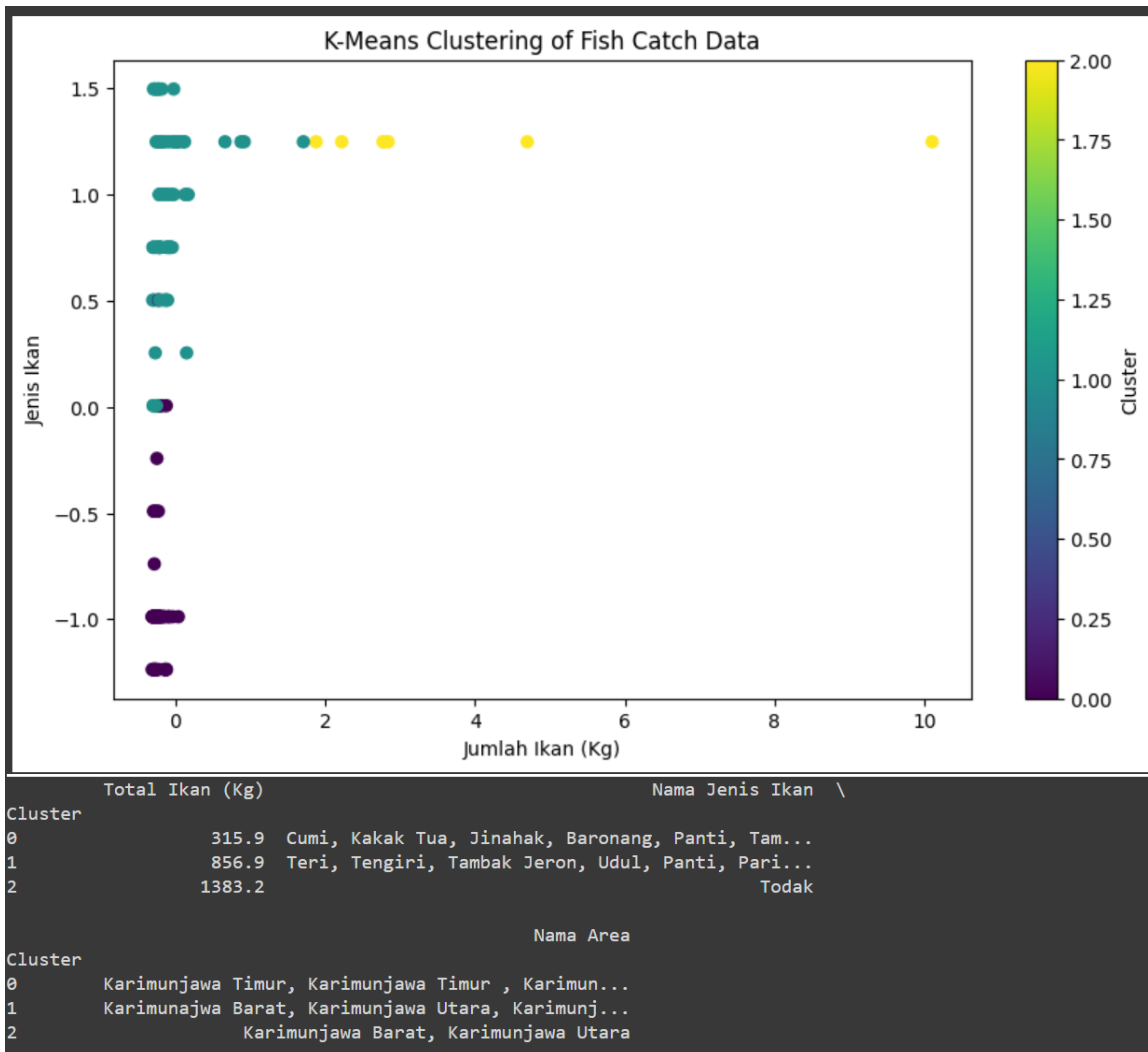
# Menampilkan informasi cluster dengan nama-nama jenis ikan dan area penangkapan
data_cleaned['Jenis Ikan'] = data['Jenis Ikan']
data_cleaned['Area Penangkapan'] = data['Area Penangkapan']

cluster_info = data_cleaned.groupby('Cluster').agg({
    'Jumlah Ikan (Kg)': 'sum',
    'Jenis Ikan': lambda x: ', '.join(x.unique()),
    'Area Penangkapan': lambda x: ', '.join(x.unique())
}).rename(columns={
    'Jumlah Ikan (Kg)': 'Total Ikan (Kg)',
    'Jenis Ikan': 'Nama Jenis Ikan',
    'Area Penangkapan': 'Nama Area'
})

print(cluster_info)

```

Gambar 6: Program Visualisasi



Gambar 7: Visualisasi Data

```

# Importing necessary libraries
from sklearn.metrics import silhouette_score, davies_bouldin_score
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Assuming data_cleaned and features_scaled are already defined
# Calculate Silhouette Score
silhouette_avg = silhouette_score(features_scaled, data_cleaned['Cluster'])
print(f'Silhouette Score: {silhouette_avg}')

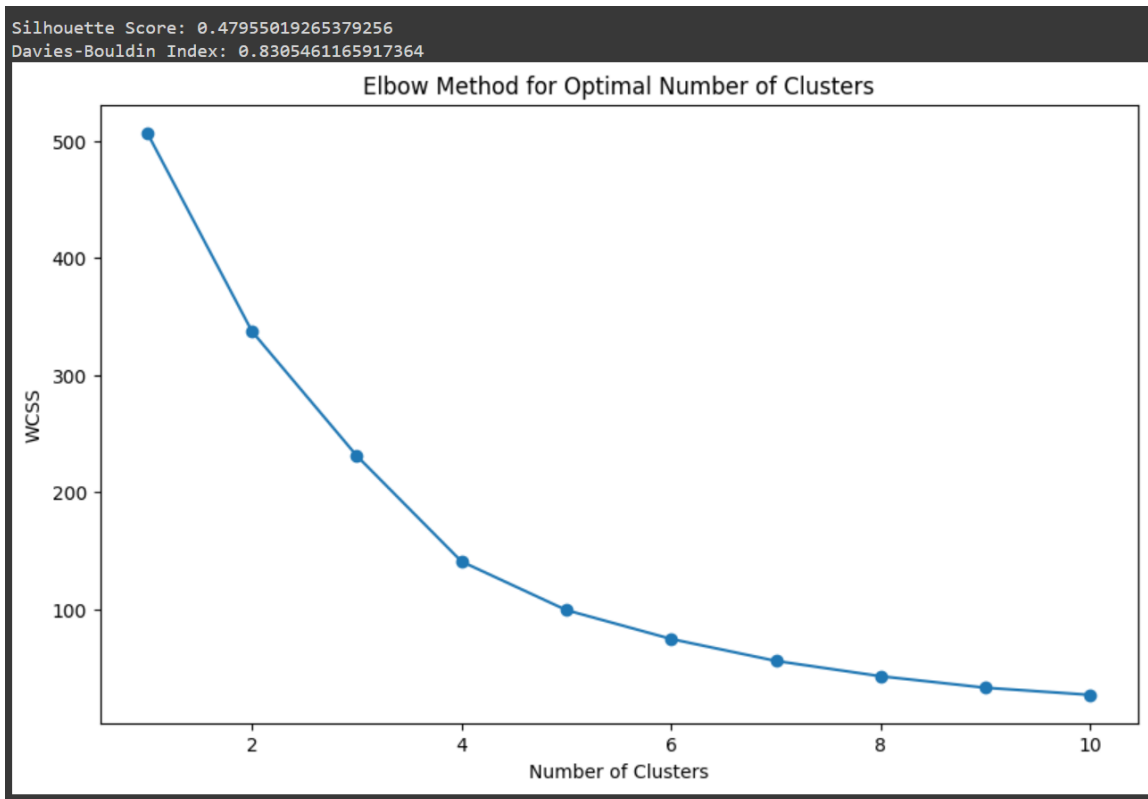
# Calculate Davies-Bouldin Index
davies_bouldin_avg = davies_bouldin_score(features_scaled, data_cleaned['Cluster'])
print(f'Davies-Bouldin Index: {davies_bouldin_avg}')

# Elbow Method for determining the optimal number of clusters
wcss = [] # Within-cluster sum of square
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=0)
    kmeans.fit(features_scaled)
    wcss.append(kmeans.inertia_)

# Plotting the Elbow Method
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11), wcss, marker='o')
plt.title('Elbow Method for Optimal Number of Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.show()

```

Gambar 8: Evaluasi



Gambar 9: Hasil Evaluasi menggunakan metrix *Silhouette Score*, *Davies-Bouldin Index*, dan *Elbow Methode*

Dalam studi ini peneliti menggunakan beberapa metrix evaluasi seperti *Silhouette Score*, *Davies-Bouldin Index*, dan *Elbow Methode* untuk membantu mengetahui kualitas kluster karena *K-Means* tidak memiliki "akurasi" dalam arti pengklasifikasian seperti supervised learning. Dalam studi ini hasil dari metrix *Silhouette Score* sebesar 0.47955019265379256 mengindikasikan seberapa baik objek berada dalam kluster mereka sendiri dibandingkan dengan kluster lain. Nilai *Silhouette Score* berkisar dari -1 hingga 1, di mana nilai lebih dekat ke 1 menunjukkan kluster yang lebih padat dan terpisah. *Davies-Bouldin Index* sebesar 0.8305461165917364 mengukur seberapa baik kluster terpisah satu sama lain. Nilai yang lebih rendah menunjukkan kluster yang lebih baik dibedakan. *Metode Elbow* membantu menentukan jumlah kluster optimal dengan melihat titik di mana penurunan varians di dalam kluster berkurang secara signifikan (elbow point).

4.2. Pembahasan

Studi ini menerapkan algoritma *K-Means* untuk mengklusterisasi data hasil tangkapan ikan di Karimunjawa, dengan tujuan untuk mengidentifikasi jenis ikan yang paling banyak ditangkap serta area penangkapan yang produktif. Data ini memberikan wawasan yang berharga dalam pengelolaan sumber daya perikanan dan pemahaman lebih dalam terhadap dinamika ekosistem perairan Karimunjawa. Algoritma *K-Means* dipilih karena kemampuannya dalam mengelompokkan data berdasarkan pola yang tidak diketahui sebelumnya. Dalam konteks ini, analisis dimulai dengan pengumpulan data tangkapan ikan yang mencakup berbagai jenis seperti Cumi, Kakak Tua, Jinahak, Baronang, Panti, Tambak Jeron, Teri, Tengiri, Udul, Pari, dan Todak. Data ini diolah untuk memisahkan jenis-jenis ikan ke dalam kluster yang mewakili kelompok dengan karakteristik tangkapan yang serupa. Dari hasil analisis, ditemukan bahwa algoritma *K-Means* berhasil mengelompokkan jenis ikan menjadi tiga kluster utama berdasarkan berat tangkapan: Kluster 0 (315.9 kg) terdiri dari jenis ikan seperti Cumi, Kakak Tua, Jinahak, Baronang, Panti, dan Tambak Jeron. Kluster ini menunjukkan pola tangkapan yang serupa untuk jenis-jenis ikan ini di seluruh area penangkapan. Kluster 1 (856.9 kg) mencakup Teri, Tengiri, Tambak Jeron, Udul, Panti, dan Pari. Jenis ikan dalam kluster ini cenderung lebih banyak ditangkap di area-area

tertentu yang memiliki karakteristik ekologis atau kondisi perairan yang mendukung keberadaan mereka. Klaster 2 (1383.2 kg) didominasi oleh Todak, menunjukkan bahwa spesies ini cenderung memiliki distribusi dan pola tangkapan yang berbeda dibandingkan dengan jenis ikan lainnya di Karimunjawa. Selain mengklasterisasi jenis ikan, studi ini juga menganalisis area penangkapan yang produktif. Berdasarkan hasil *K-Means*, area penangkapan dibagi menjadi: Karimunjawa Timur (Klaster 0), Karimunjawa Barat (Klaster 1), Karimunjawa Utara (Klaster 2). Pembagian ini membantu mengidentifikasi area-area yang memperlihatkan produktivitas yang tinggi dalam hal tangkapan ikan tertentu. Analisis ini penting untuk mengarahkan kebijakan pengelolaan perikanan yang berkelanjutan, memastikan pemanfaatan sumber daya alam yang efektif dan menjaga keseimbangan ekosistem laut. Secara keseluruhan, implementasi algoritma *K-Means* dalam studi ini memberikan pemahaman yang mendalam tentang struktur tangkapan ikan di Karimunjawa. Hasilnya tidak hanya mengidentifikasi pola tangkapan berdasarkan jenis ikan, tetapi juga memberikan wawasan tentang distribusi area penangkapan yang berkontribusi terhadap produktivitas perikanan di wilayah tersebut. Informasi ini diharapkan dapat mendukung keputusan yang lebih baik dalam pengelolaan dan pelestarian sumber daya perikanan Karimunjawa untuk masa depan yang berkelanjutan.

KESIMPULAN

Studi ini berhasil mengimplementasikan algoritma K-Means Clustering menggunakan Google Colab dan bahasa pemrograman Python untuk menganalisis data tangkapan ikan di Karimunjawa. Tujuan utama studi adalah untuk mengklasterisasi jenis ikan yang paling banyak ditangkap serta mengidentifikasi area penangkapan yang produktif di wilayah tersebut. Dengan memanfaatkan dataset berat tangkapan ikan yang mencakup berbagai spesies seperti Cumi, Kakak Tua, Jinahak, Baronang, Panti, Tambak Jeron, Teri, Tengiri, Udul, Pari, dan Todak, algoritma K-Means berhasil mengelompokkan data ke dalam tiga klaster utama: Klaster 0 (315.9 kg) mencakup Cumi, Kakak Tua, Jinahak, Baronang, Panti, dan Tambak Jeron, menunjukkan pola tangkapan yang serupa di beberapa lokasi penangkapan di Karimunjawa Timur. Klaster 1 (856.9 kg) terdiri dari Teri, Tengiri, Tambak Jeron, Udul, Panti, dan Pari, dengan preferensi tangkapan yang lebih dominan di area Karimunjawa Barat. Klaster 2 (1383.2 kg) didominasi oleh Todak, menunjukkan distribusi yang berbeda dan tangkapan yang lebih banyak di Karimunjawa Utara. Hasil evaluasi menggunakan metrik juga memberikan wawasan yang berarti. Silhouette Score yang positif (0.48) menunjukkan bahwa klaster yang dihasilkan cenderung terpisah dengan baik, meskipun ada ruang untuk perbaikan lebih lanjut dalam pemisahan klaster. Davies-Bouldin Index yang rendah (0.83) mengindikasikan bahwa klaster cukup terpisah satu sama lain, meskipun ada overlap yang potensial antara beberapa klaster. Metode Elbow digunakan untuk menentukan jumlah klaster optimal, memberikan panduan dalam pemilihan konfigurasi yang tepat untuk analisis klaster. Kesimpulannya, implementasi K-Means Clustering dalam studi ini tidak hanya berhasil mengelompokkan jenis ikan berdasarkan pola tangkapan dan area penangkapan yang produktif di Karimunjawa, tetapi juga memberikan landasan penting bagi pengelolaan perikanan yang berkelanjutan di wilayah tersebut. Informasi ini dapat digunakan untuk mengoptimalkan strategi penangkapan ikan dan menjaga keberlanjutan ekosistem perairan, sesuai dengan tujuan perlindungan sumber daya alam yang krusial bagi masa depan Karimunjawa.

REFERENSI

- Apriyani, Putri, Arif Rinaldi Dikananda, and Irfan Ali. 2023. "Penerapan Algoritma K-Means Dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi." *Hello World Jurnal Ilmu Komputer* 2(1):20–33. doi: 10.56211/helloworld.v2i1.230.
- Ejimofor, Ihekeremma A. U., and Obi O. R. Okonkwo. 2021. "Development of a Knowledge Discovery System in Big Data Mining Environment." *International Research Journal of Innovations in Engineering and Technology* 05(08):65–70. doi: 10.47001/irjiet/2021.508011.
- Elda, Yusma, Sarjon Defit, Yuhandri Yunus, and Raemon Syaljumairi. 2021. "Klasterisasi Penempatan Siswa Yang Optimal Untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan

- K-Means.” *Jurnal Informasi Dan Teknologi* 103–8. doi: 10.37034/jidt.v3i3.130.
- Hayati, Umi. 2023. *CLUSTERING PRODUK EKSPOR INDONESIA BERDASARKAN TINGKAT PERMINTAAN MENGGUNAKAN METODE K-MEANS TAHUN 2020-2022*. Vol. 7.
- Jo, Jongkwon, Seungha Jung, Joongyang Park, Youngsoon Kim, and Mingon Kang. 2022. “Hi-LASSO: High-Performance Python and Apache Spark Packages for Feature Selection with High-Dimensional Data.” *PLoS ONE* 17(12 December):1–8. doi: 10.1371/journal.pone.0278570.
- Julianti Hablum, Rofika, Amal Khairan, Ji Jati Metro, and Kota Ternate Selatan. 2019. *JIKO (Jurnal Informatika Dan Komputer) Ternate CLUSTERING HASIL TANGKAP IKAN DI PELABUHAN PERIKANAN NUSANTARA (PPN) TERNATE MENGGUNAKAN ALGORITMA K-MEANS*.
- Laccetti, Giuliano, Marco Lapegna, Valeria Mele, Diego Romano, and Lukasz Szustak. 2020. “Performance Enhancement of a Dynamic K-Means Algorithm through a Parallel Adaptive Strategy on Multicore CPUs.” *Journal of Parallel and Distributed Computing* 145:34–41. doi: 10.1016/j.jpdc.2020.06.010.
- Luan, Lan. 2020. “Optimized Implementation of K_means Clustering Algorithm on Reconfigurable System.” Pp. 856–59 in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*. IEEE.
- Nabila, Zulfa, Auliya Rahman Isnain, and Zaenal Abidin. 2021. “ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS.” *Jurnal Teknologi Dan Sistem Informasi (JTSI)* 2(2):100.
- Pilnenskiy, Nikita, and Ivan Smetannikov. 2020. “Feature Selection Algorithms as One of the Python Data Analytical Tools.” *Future Internet* 12(3). doi: 10.3390/fi12030054.
- Prasetyo, Rangga Bayu, Yosep Agus Pranoto, and Renaldi Primaswara Prasetya. 2023. *IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PENYAKIT PASIEN RAWAT JALAN PADA KLINIK DR. ATIRAH DESA SIOYONG, SULTENG*. Vol. 7.
- Virgo, Ismail, Sarjon Defit, and Y. Yuhandri. 2020. “Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering.” *Jurnal Sistim Informasi Dan Teknologi* 23–28. doi: 10.37034/jsisfotek.v2i1.17.
- Wardono, Budi, Rizky Muhartono, Yayan Hikmayani, Tenny Apriliani, and Hikmah Hikmah. 2019. “ANALISIS PROSPEKTIF PERAN AKTOR DALAM STRATEGI FORMULASI PEMBANGUNAN PERIKANAN DI KABUPATEN NATUNA.” *Jurnal Sosial Ekonomi Kelautan Dan Perikanan* 14(2):179. doi: 10.15578/jsekp.v14i2.8241.
- Xie, Ting, Ruihua Liu, and Zhengyuan Wei. 2020. “Improvement of the Fast Clustering Algorithm Improved by K-Means in the Big Data.” *Applied Mathematics and Nonlinear Sciences* 5(1):1–10. doi: 10.2478/amns.2020.1.00001.