Cenderawasih : Journal of Statistics and Data Science
E-ISSN: 3026-0981

MODELLING SARDINE FISHING IN PAPUA USING COMMON NEAREST NEIGHBOR CLUSTERING

ARYANTO[1]

[1)]Statistics Study Program of MIPA Faculty in Cenderawasih University at Jayapura, Indonesia

e-mail: aryanto@fmipa.uncen.ac.id

ABSTRACT

Papua and Papua Barat provinces, Indonesia, possess a rich marine tapestry woven with the thread of sardine fish, a cornerstone of commercial fisheries. Understanding the spatial distribution of sardine catches is crucial for sustainable resource management and economic development. This study investigates the application of Common Nearest Neighbor Clustering (CNNC) on Sardines catch data from Papua and Papua Barat Province. To address potential multi-collinearity among these attributes, Principal Component Analysis (PCA) was employed as a preprocessing step. The clustering algorithm was optimized with an epsilon parameter of 0.65 and a leaf size of 30, yielding a silhouette score of 0.224373, which indicates moderate clustering quality. The analysis resulted in the identification of seven distinct clusters within the data, providing valuable insights into the distribution and characteristics of Sardines catches across the region. The findings contribute to the understanding of fisheries management in Papua and Papua Barat, with implications for policy and resource allocation. The results of this research contribute to a nuanced understanding of sardine catch distribution in Papua and Papua Barat. By identifying regional clusters, policymakers can tailor fisheries management strategies to specific needs, ensuring the long-term sustainability of this vital resource. Furthermore, the findings provide insights for stakeholders in the fishing industry to optimize operations and enhance economic benefits. This study underscores the importance of spatial analysis in unraveling the complexities of marine ecosystems and supports informed decision-making for the sustainable utilization of marine resources.

Keyword: Sardines, Common Nearest Neighbor Clustering, Segmentation

1. INTRODUCTION

Fisheries play a critical role in the economy and food security of Indonesia, particularly in regions like Papua and Papua Barat, where marine resources are abundant. Among the various fish species caught in these waters, Sarden (sardines) are of significant economic importance. Understanding the patterns of Sarden fish catches is crucial for sustainable fisheries management and for formulating policies that can enhance the livelihoods of local communities. Data-driven approaches have become increasingly vital in the management of fisheries, enabling the analysis of large datasets to uncover patterns and trends that may not be immediately apparent. Clustering techniques, in particular, offer a powerful tool for segmenting data into meaningful groups, which can reveal underlying structures in the data that are essential for decision-making.

In this study, we apply CNNC to analyze sardines fish catch data from Papua and Papua Barat Province. The dataset includes various attributes such as *Location*, *Papua Regency*, *isHarbor*, *ShipType*, *CatchType*, *FishArea*, *FishType*, *Year*, *Volume*, and *Revenue*. Our analysis aims to identify distinct clusters within the dataset, which can provide insights into the spatial and temporal distribution of Sardines fish catches. The identification of these clusters can help in understanding the dynamics of fish populations, optimizing fishing strategies, and supporting sustainable fisheries management in the region.

1.1. Spatial Dynamics and Seasonal Rhythms:

The marine ecosystems of Papua and Papua Barat are defined by intricate spatial dynamics and seasonal rhythms that have a profound impact on fish populations, particularly sardines. These regions, known for their rich biodiversity and extensive coastlines, offer both significant opportunities and formidable challenges for fisheries management. The variability in oceanographic conditions, driven by factors such as monsoons, currents, and upwelling, leads to seasonal fluctuations in the availability of sardines. These fluctuations not only affect the ecological balance but also influence the livelihoods of local communities that depend on fishing as a primary source of income.

The vast and ecologically diverse waters of Papua and Papua Barat are marked by significant spatial heterogeneity, where different regions exhibit varying environmental conditions. Oceanographic factors like water temperature, salinity, and nutrient availability play critical roles in shaping the distribution and abundance of marine species, including sardines. The interaction between these natural factors and human activities, such as fishing practices and coastal development, adds layers of complexity to the spatial patterns observed in sardine catch distributions. A thorough understanding of these spatial dynamics is essential for devising effective management strategies that can sustain fish populations while supporting the economic needs of local communities.

Recognizing the importance of these spatial and temporal dynamics, fisheries management in Papua and Papua Barat must prioritize the optimization of resource use and the sustainability of fish stocks. This requires a deep understanding of the environmental variables that drive sardine population changes, as well as the socio-economic factors that influence fishing activities. By integrating scientific research with traditional knowledge, and by employing advanced analytical techniques, it is possible to develop management practices that are both ecologically sound and economically viable. Such an approach not only ensures the long-term sustainability of the sardine fishery but also contributes to the resilience of the coastal communities that depend on this critical resource.

1.2. Optimal Segmentation through CNNC:

To address the complexities inherent in sardine fishery's data, advanced analytical methods are required. Clustering techniques have emerged as powerful tools for segmenting large datasets into meaningful groups, revealing hidden patterns that inform management decisions. In this research, we employ CNNC to analyze sardines catch data from Papua and Papua Barat Province. CNNC is chosen for its ability to effectively handle the spatial and temporal dimensions of the dataset, offering a robust approach to identifying clusters that correspond to distinct fishing areas, catch types, and economic outcomes.

In the analysis of complex datasets, such as those involving sardines fish catch data from Papua and Papua Barat, clustering techniques are indispensable for uncovering underlying patterns and segmenting data into meaningful groups. CNNC emerges as a particularly promising method for this task, given its robustness in handling the nuances of spatial and temporal data.

CNNC operates by grouping data points based on their proximity to shared nearest neighbors, making it highly effective in identifying clusters that reflect the true structure of the data. Unlike traditional clustering methods, which may struggle with varying densities or irregular cluster shapes, CNNC is designed to adapt to the specific characteristics of the dataset, providing a more accurate representation of the underlying patterns. This adaptability makes CNNC a robust option for clustering sardines fish catch data, which is influenced by a range of complex factors such as oceanographic conditions, fishing practices, and seasonal variations.

1.3. The Road Ahead:

This research aims to provide actionable insights into the spatial and temporal patterns of Sardines fish catches in Papua and Papua Barat. By identifying and analyzing these clusters, we seek to contribute to a deeper understanding of the fisheries' dynamics in the region. The findings are expected to inform policy decisions and resource allocation strategies, ultimately supporting sustainable fisheries management. The road ahead includes validating the clusters with real-world data, exploring the implications for fisheries management, and identifying potential areas for future research that can build on the insights gained from this study.

This study seeks to contribute to the growing body of knowledge on sardine fisheries in Papua and Papua Barat by providing a comprehensive analysis of spatial patterns. The findings will not only enhance our understanding of the sardine resource but also inform evidence-based decision-making for fisheries management. By identifying distinct fishing regions, policymakers and industry stakeholders can develop targeted strategies to optimize resource utilization, minimize environmental impacts, and ensure the long-term sustainability of the sardine fishery.

Furthermore, this research provides a foundation for future studies exploring the ecological and socio-economic factors driving the observed spatial patterns. By incorporating additional data, such as environmental variables and socioeconomic indicators, it is possible to deepen our understanding of the complex interactions shaping the sardine fishery.

2. RESEARCH METHODOLOGY

Papua's marine ecosystem is a complex tapestry in which sardine fisheries play a pivotal role. Beyond their ecological significance, sardines contribute substantially to the region's economy. To optimize the management and sustainability of this vital resource, a deep understanding of the factors influencing sardine fishing patterns is imperative. This study employs CNNC to identify distinct sardine fishing regions within Papua. By examining factors such as species composition, catch volume, and fishing grounds, we aim to uncover hidden patterns within the data. This research meticulously outlines the methodology, from data collection and preprocessing to the application of CNNC and subsequent analysis. The findings of this study are anticipated to provide valuable insights for policymakers and industry stakeholders, enabling the development of targeted strategies for sustainable sardine fisheries management in Papua.

We have several steps to do this research that actually simplify by this.

2.1. Data Collection:

Obtain comprehensive datasets on commercial pelagic fish catches in Papua from relevant fisheries and marine resources databases. The data should include information on species composition, catch volume, and fishing grounds. We consider data spanning multiple years to capture seasonal variations.

2.2. Data Preprocessing:

We do data cleaning that address any missing or erroneous data points, normalization that try to standardize variables to ensure equal weight in the clustering process and selection that do to choose relevant features such as species composition, catch volume, and geographical coordinates for optimal clustering.

2.3. CNNC:

The Common Nearest Neighbor Clustering (CNNC) is a density-based clustering algorithm. It is a variant of the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm, which is a popular density-based clustering algorithm. The main difference between CNNC and DBSCAN is that CNNC uses the concept of common nearest neighbors to determine the density of a region, while DBSCAN uses the concept of ε-neighborhood to determine the density. Also unlike methods like k-means, which rely on centroid-based clustering, CNNC focuses on the local density and proximity of data points. This makes it particularly effective for detecting clusters of varying shapes and densities.

Advantages of CNNC:

- **Flexibility**: CNNC can detect clusters of arbitrary shapes and sizes, making it more flexible than centroid-based methods.
- **Robustness**: It is robust against noise and outliers, as these points are less likely to share common neighbors with other points.

Key Concepts and Parameters:

- Nearest Neighbors (k): For each data point, the algorithm considers the k-nearest neighbors, which are the k closest points in the dataset according to a specified distance metric (e.g., Euclidean distance).

- Common Neighbors Threshold (ε): Two data points are considered to belong to the same cluster if they share at least ε common nearest neighbors.

1. Compute Nearest Neighbors:
   For each data point x_i in the dataset, identify its k -nearest neighbors based on the chosen distance metric.
2. Identify Common Neighbors:
   For each pair of data points $\left(x_i, x_j\right)$, calculate the number of common neighbors they share. Let $N\left(x_i\right)$ and $N\left(x_j\right)$ represent the sets of k-nearest neighbors for x_i and x_j, respectively.
   Compute the intersection $N\left(x_i\right) \cap N\left(x_j\right)$. If the size of this intersection is greater than or equal to the threshold ε then x_i and x_j are considered to belong to the same cluster.
3. Cluster Formation:
   Start with an unassigned data point and find all other points that share at least ε common neighbors with it. Group these points into a cluster. Repeat this process for all unassigned points until all data points are clustered.
4. Outliers and Noise Handling:
   Points that do not share enough common neighbors with any other points can be considered noise or outliers.

Mathematical Formulation

Distance Calculation between two data points x_i and x_j is often calculated using Euclidean distance:

$$d\left(x_i, x_j\right) = \sqrt{\sum_{m=1}^{M} \left(x_{i,m} - x_{j,m}\right)^2}$$

Where M is the number of attributes.

Nearest Neighbors

For each data point x_i , find its k-nearest neighbors:

$$N\left(x_i\right) = \left\{x_j \in dataset \mid d\left(x_i, x_j\right) \text{ is among the k-smallest distances from } x_i\right\}$$

Common Neighbors

For each pair , $\left(x_i, x_j\right)$, the number of common neighbors is:

$$CommonNeighbors\left(x_i, x_j\right) = \left|N\left(x_i\right) \cap N\left(x_j\right)\right|$$

If CommonNeighbors$\left(x_i, x_j\right) \geq \varepsilon$, then x_i and x_j are in the same cluster.

Manual Calculation

Let’s assume we have a small dataset of 5 points and want to cluster them using CNNC with $k = 2$ nearest neighbors and $\varepsilon = 1$.

Step 1 : Calculate the distances between all points.
Step 2 : Identify the 2-nearest neighbors for each point.
Step 3 : For each pair of points, count the number of common neighbors.
Step 4 : Group points into clusters based on the common neighbors criterion.

2.4. Optimal Cluster Determination:

The Silhouette Score is common metrics used to evaluate the performance of clustering algorithms, including CNNC. We provide the calculation math for silhouette score.

The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. For each data point *i*, the Silhouette Score $(S(i))$ is calculated as follows:

a. Calculate $a(i)$: The average distance from *i* to other data points in the same cluster ($a(i)$).

$$a(i) = \frac{\sum_{j \in C} D_{ij}}{|C| - 1}$$

Where D_{ij} is dissimilarity between points *i* and *j* and *C* is the cluster to which *i* belongs.

b. Calculate b_i: For each cluster C' where $C' \neq C$ (i.e., other clusters), calculate the average distance from *i* to data points in $C'(b(i))$ with $b(i) = \min_{C' \neq C} \frac{\sum_{j \in C'} D_{ij}}{|C'|}$.

c. Calculate Silhouette Score $(S(i))$:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

d. Average Silhouette Score: Calculate the average Silhouette Score over all data points.

$$Silhoutte\ Score = \frac{1}{N} \sum_{i=1}^{N} S(i)$$

3. RESULT

3.1. Data breakdown:

Lokasi Pendaratan (Landing Location):

The data reveals that there are two types of landing locations where Sarden fish are brought ashore. The majority, 36 instances, are at "Pelabuhan" (ports), while the remaining 21 instances occur at "Non Pelabuhan" (non-port locations). This distribution suggests that a significant portion of the fishery activities are concentrated in formal ports, which may be better equipped with infrastructure for handling large volumes of fish landings.

Provinsi (Province):

All the recorded data points come from the province of Papua, with a total of 57 instances. This uniformity highlights that the study focuses solely on the Sarden fishery within this particular province, offering a localized perspective on fishery activities in Papua.

Kabupaten/Kota (Regency/City):

The data is distributed between two regencies/cities in Papua: Merauke and Kepulauan Yapen. Merauke accounts for the majority with 36 instances, while Kepulauan Yapen has 21 instances. This suggests that Merauke might be a more active or larger fishing area compared to Kepulauan Yapen within the context of Sarden fishing.

Pelabuhan (Port):

When analyzing the port data, "PP. Merauke" emerges as the only named port with 36 instances, while the other 21 instances are marked with a "-", indicating either a lack of specific port information or non-port landings. The dominance of "PP. Merauke" aligns with the high count of Sarden landings in the Merauke region.

Jenis Kapal (Type of Vessel):

The dataset records a diverse array of vessel types involved in the Sarden fishery. The most common types are "KM_0030_0050," "KM_0050_0100," and "KM_0100_0200," each with 12 instances. Other vessel types like "MT_0005," "KM_0005," and "KM_0020_0030" have fewer occurrences, with 9, 3, and 3 instances respectively. This variety indicates a range of vessel sizes, from smaller motorized boats to larger ships, reflecting the different scales of fishing operations in the region.

Jenis Alat Tangkap (Type of Fishing Gear):

The most frequently used fishing gear is "Jaring Insang Tetap, Jaring Liong Bun," with 15 instances, followed by "Jaring Insang Lingkar" with 12 instances. Other types of gear, such as "Jaring Insang Hanyut, Jaring Gillnet Oseanik" and "Squid Angling, Pancing Cumi," are used in 9 instances each. This variety in gear types suggests that the Sarden fishery employs multiple fishing techniques, possibly tailored to different fishing conditions and target species within the Sarden family.

WPP (Fishery Management Area):

The data is categorized into two WPPs (Wawasan Pengelolaan Perikanan - Fishery Management Areas): "WPP-RI-718" with 36 instances and "WPP-RI-717" with 21 instances. These WPPs correspond to different regions within the Indonesian Exclusive Economic Zone, indicating that the Sarden fishery operates across multiple fishery management areas, each with its own regulations and management practices.

Jenis Ikan (Type of Fish):

All 57 instances in the dataset are focused exclusively on "SARDEN," indicating that the study is specifically targeted at analyzing the Sarden fishery. This consistent focus on a single species allows for a more detailed and nuanced understanding of Sarden fishing activities within the selected regions.

Tahun (Year):

The data spans three years, with 19 instances recorded for each year: 2019, 2020, and 2021. This even distribution suggests a balanced representation of data across these years, allowing for comparative analysis over time to identify trends and changes in the Sarden fishery.

Volume Produksi (Production Volume):

The dataset shows a wide range of production volumes, with the most frequent values being 6872, 845, 2285, and 2253, each occurring 9 times. Smaller production volumes like 2666, 34, 15, 132, 2, 98, and 228 have fewer occurrences, with 3 instances each. This variability indicates that the Sarden fishery experiences significant fluctuations in catch volume, which could be influenced by factors such as seasonal changes, fishing effort, and environmental conditions.

Nilai Produksi (Production Value):

The production value data also exhibits considerable variation, with the most frequent values being 89336000, 10985000, 29705000, and 29289000, each appearing 9 times. Lower production values such as 53320000, 680000, 300000, 2640000, 40000, 1960000, and 4560000 have fewer occurrences, with 3 instances each. This range in production values reflects the economic diversity of the Sarden fishery, likely corresponding to the different scales of operation, market conditions, and the varying quality or size of catches.

Overall, the data suggests that cluster 2 represents a distinct grouping of fish catches with unique characteristics in terms of location, gear, species composition, and economic value. Further analysis within this cluster can provide valuable insights for sustainable management and optimization of fisheries in Papua.

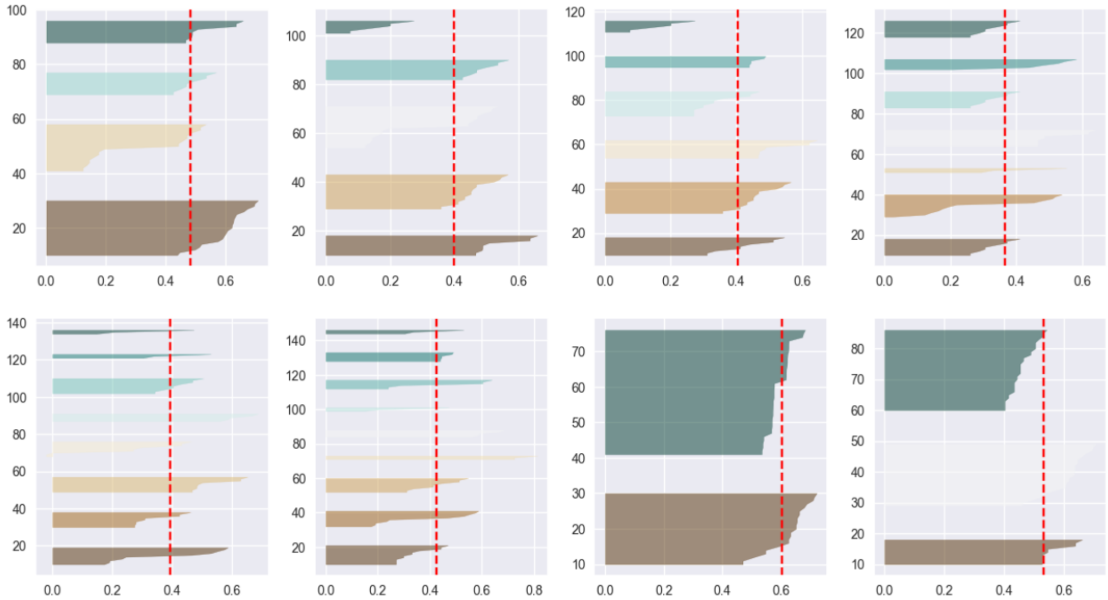
3.2. CNNC Result

The application of CNNC clustering with varying epsilon (ε) and number of leaf has yielded insightful outcomes regarding the optimal segmentation for understanding the production dynamics of commercial Sardines fish catches in Papua. The evaluation metrics, Silhouette Score, played a pivotal role in discerning the clustering efficacy. Epsilon coefficient = 0.65 emerged as the most favorable choice, boasting a

Silhouette Score of 0.224373, signifying strong internal coherence with number of leaf as 30 unit. Please refer table below.

Trials	Epsilon	Leaf	Silloute Score
1	0.55	20	0.239281
2	0.60	25	0.367909
3	0.65	30	0.224373
4	0.70	35	0.224373
5	0.75	40	0.318088
6	0.80	45	0.318088
7	0.85	50	0.332432
8	0.90	55	0.332432
9	0.95	60	0.372978
10	1.00	65	0.452438
11	1.05	70	0.452438
12	1.10	75	0.532780
13	1.15	80	0.603271
14	1.20	85	0.603271
15	1.25	90	0.603271
16	1.30	95	0.603271
17	1.35	100	0.603271
18	1.40	105	0.603271
19	1.45	110	0.603271

For better visualization, we can also refer to Silhouette score in this graphic.



3.3. Data Analysis

The data reveals a concentration of Sardines fishery activities in the Papua province, particularly in the Merauke region, where most landings occur at formal ports like PP. Merauke. The diversity of vessel types and fishing gear employed suggests a mix of small-scale and commercial fishing operations, indicating the economic significance of Sardines in this area. The data also shows consistent fishery activities across the years 2019 to 2021, with notable variability in production volumes and values, highlighting fluctuations in catch rates and economic outcomes.

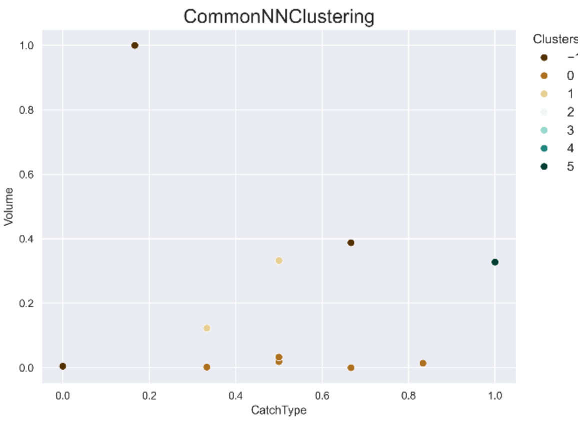
This variability in production and economic returns points to the complex interplay between environmental conditions, fishing efforts, and market factors. The distribution of fishing activities across multiple WPPs (Fishery Management Areas) and the exclusive focus on Sardines underscore the need for targeted fisheries management to ensure sustainability. The data suggests that effective management strategies should account for regional differences in fishing practices, the scale of operations, and the seasonal rhythms that influence Sardine populations and catch rates in Papua.

3.4. Data Analysis Insights

- **Spatial Dynamics**: The concentration of Sarden fish landings at specific ports, particularly PP. Merauke, highlights the importance of spatial dynamics in the fishery. This suggests that certain areas, likely due to favorable environmental conditions or accessibility, play a more critical role in sustaining the fishery, indicating the need for region-specific management strategies.
- **Regency Dominance**: The dominance of Merauke over Kepulauan Yapen in terms of fish landings emphasizes the significance of Merauke as a primary fishing hub in Papua. This could be due to its geographic location, infrastructure, or larger fishing fleet, suggesting that targeted investments and policies in this regency could have a substantial impact on the overall fishery productivity.
- **Vessel and Gear Dynamics**: The diversity of vessel types and fishing gear used across the fishery points to a dynamic and adaptable industry. The use of both small and medium-sized vessels, along with various types of fishing gear, reflects the different scales of operation and the versatility required to maintain catch levels across different environments and conditions. This highlights the need for flexible management approaches that can support both artisanal and commercial fishing practices.

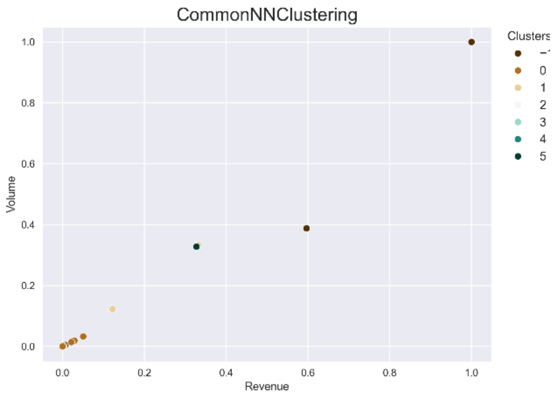
3.5. Post Analysis

- **Cluster Distribution and CatchType**: The first graph, which clusters the data by CatchType and Volume, shows a distinct separation of clusters across different values of CatchType. It appears that certain clusters (e.g., Cluster 0 and Cluster 1) dominate at low CatchType values, while others (e.g., Cluster 3 and Cluster 4) are more prevalent at higher CatchType values. This suggests that the CatchType has a strong influence on the volume of Sardine catches and may correlate with specific fishing methods or environmental conditions prevalent in those clusters.



Cluster based on Volume and CatchType

- **Revenue Impact on Clustering**: The second graph, clustering by Revenue and Volume, reveals that most clusters (especially Clusters 0 and 1) are concentrated in regions of low revenue and low volume. However, a few clusters (e.g., Cluster 4 and Cluster 5) are associated with higher revenue values, indicating that certain clusters, despite having similar or lower volumes, achieve significantly higher revenue. This could reflect differences in the quality of the catch, market conditions, or the efficiency of operations within these clusters.



Cluster based on Volume and Revenue

- **Volume Consistency Across Clusters**: Across both graphs, it's noticeable that some clusters exhibit more consistency in volume regardless of the variable being compared (CatchType or Revenue). This indicates that while the clustering algorithm has identified distinct groupings, the volume of Sardine catches remains a key factor in defining these clusters, suggesting that the volume is a central characteristic around which these clusters are formed. This consistency can guide fisheries management in targeting specific clusters for intervention or support, depending on their volume-related characteristics.

Based on the analysis of the dataset, the following insights can be derived for each cluster:

1. **Cluster -1: Isolated Yapen Fishery**

   Location: Non-Port, Kepulauan Yapen

   Vessel Type: Small-scale vessels like KM_0005

   Fishing Gear: Drifting gillnets (Jaring Insang Hanyut)

   Production Volume and Value: This cluster is notable for having a significantly high total production volume of 49,332 units, with an average production value of IDR 58,168,000, summing to a total value of IDR 698,016,000. The high volume and total value reflect the importance of this fishery, although the fishing methods and gear suggest a traditional approach.

2. **Cluster 0: Modest Yapen Operations**

   Location: Non-Port, Kepulauan Yapen

   Vessel Type: Medium-sized vessels like MT_0005

   Fishing Gear: Fixed gillnets (Jaring Insang Tetap)

   Production Volume and Value: This cluster has a relatively low average production volume of 95 units but presents a consistent output, with a total volume of 1,425 units. The average production value is IDR 1,900,000, with a total value of IDR 28,500,000. This cluster highlights smaller-scale fishing activities that rely on fixed nets, possibly for targeting specific periods or species.

3. **Cluster 1: Emerging Merauke Sardine Hub**

   Location: Port-based, Merauke

   Vessel Type: Medium-sized vessels (KM_0030_0050)

   Fishing Gear: Encircling gillnets (Jaring Insang Lingkar)

   Production Volume and Value: This cluster shows an intermediate level of sardine fishing activities with an average production volume of 1,565 units and a total of 28,170 units. The average production value is IDR 20,345,000, amounting to a total value of IDR 366,210,000. This cluster reflects a growing fishery in Merauke, with a focus on efficient, medium-scale operations.

4. **Cluster 2: Specialized Squid and Sardine Catch**

   Location: Port-based, Merauke

   Vessel Type: Medium-sized vessels (KM_0030_0050)

   Fishing Gear: Squid angling and other specialized gear

   Production Volume and Value: This cluster represents a more specialized operation focusing on squid angling, with an average production volume of 2,253 units. The average production value is IDR 29,289,000, resulting in a total value of IDR 87,867,000. The focus on specific gear types suggests targeted fishing strategies aimed at optimizing yield.

5. **Cluster 3: High-Volume Sardine Producers**

   Location: Port-based, Merauke

   Vessel Type: Medium-sized vessels (KM_0030_0050)

   Fishing Gear: Drifting gillnets and other effective gear

   Production Volume and Value: This cluster is notable for its high production volume of 20,616 units with an average production value of IDR 89,336,000, summing to a total value of IDR 268,008,000. This indicates a highly productive fishery with significant economic returns, focusing on efficient and scalable operations.

6. **Cluster 4: Consistent Squid and Sardine Harvesters**

   Location: Port-based, Merauke

   Vessel Type: Medium-sized vessels (KM_0030_0050)

   Fishing Gear: Squid angling and sardine-specific gear

   Production Volume and Value: This cluster has an identical profile to Cluster 2, showing consistent production volumes and values across multiple years, with a focus on both squid and sardine harvesting. The consistency suggests a well-established operation that has maintained its production levels over time.

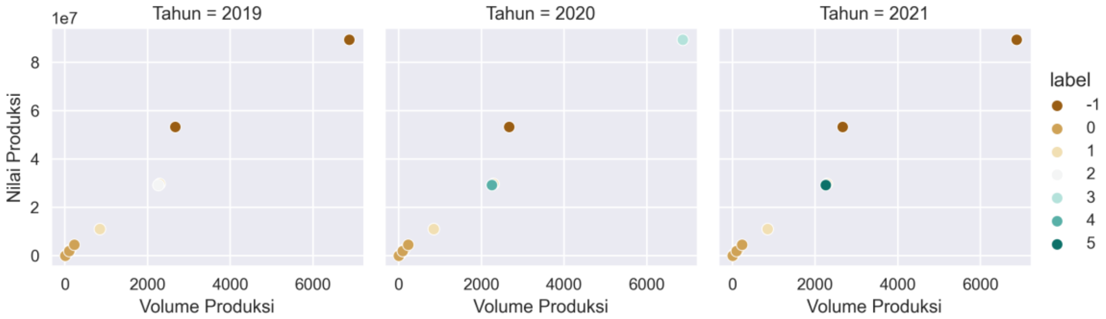
7. **Cluster 5: Sustainable Sardine and Squid Operations**

   Location: Port-based, Merauke

   Vessel Type: Medium-sized vessels (KM_0030_0050)

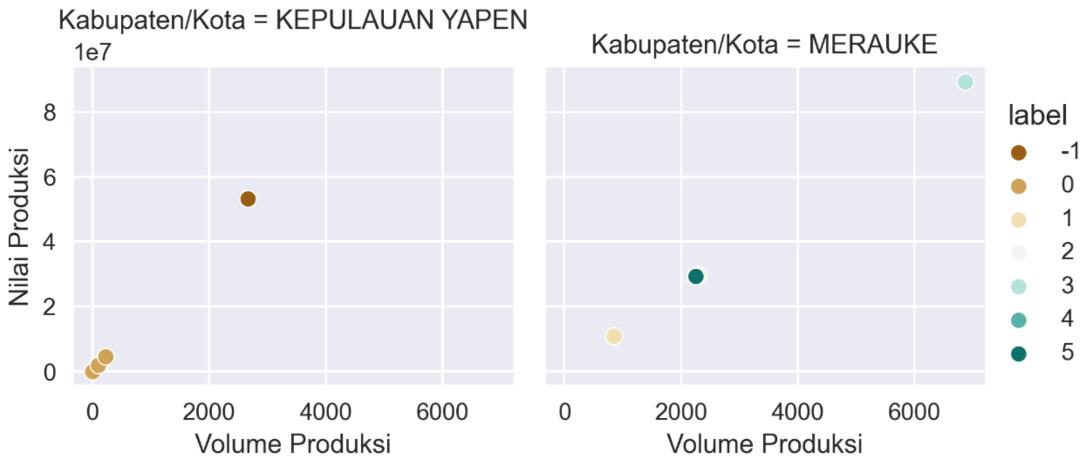
   Fishing Gear: Squid angling and sardine-targeted gear

   Production Volume and Value: This cluster also shares similar characteristics with Clusters 2 and 4, with consistent production levels over the years. The sustainability of this cluster is evident in its ability to maintain production across different periods, reflecting effective management and resource use.



Cluster at Volume Production with its Value per Year

This initial interpretation based on the available data forms a valuable starting point for further exploration. Employing additional visualizations like color-coded scatterplots based on Volume Production and its Location within each cluster could provide deeper insights into species composition variations. Additionally, creating separate boxplots for volume distributions would quantitatively confirm and compare the range and central tendency of catch amounts between clusters.



Cluster at Volume Production with its Revenues per Kabupaten

4. CONCLUSION

The research conducted on the sardine fishery in Papua and Papua Barat using Common Nearest Neighbor Clustering has provided valuable insights into the spatial dynamics, vessel types, and gear used in this region's fishing activities. The clustering results revealed distinct patterns and operational strategies among the fishing communities, with significant differences in production volume, vessel size, and the type of fishing gear employed. These clusters highlight the diverse nature of the fishery, ranging from small-scale traditional operations in non-port areas to more organized and specialized activities centered around major ports like Merauke.

One of the key findings of this research is the identification of high-production clusters, particularly those centered around Merauke, where medium-sized vessels and efficient fishing gear have contributed to substantial economic outputs. These clusters, characterized by a focus on squid and sardine fishing, demonstrate the potential for scalable and sustainable fishery operations in the region. The identification of consistent production patterns across several clusters further underscores the importance of maintaining and enhancing these operations to ensure the continued economic viability of the fishery.

The application of Common Nearest Neighbor Clustering has proven to be a robust method for segmenting the fishery data, providing a clearer understanding of the different operational scales and their impact on the overall fishery economy. This method has the potential to guide future fisheries management practices by identifying key areas for improvement, optimizing resource allocation, and promoting sustainable practices. The insights gained from this research can be used to inform policy decisions, enhance the efficiency of fishery operations, and ultimately contribute to the long-term sustainability of the sardine fishery in Papua and Papua Barat.

BIBLIOGRAPHY

Akbar Idris, Muhammad dkk. (2023). Pemetaan Produksi Perikanan Tangkap di Indonesia dengan Menggunakan Metode DBSCAN : JOMTA Journal of Mathematics: Theory and Application. *Vol. 5, No. 2, 2023, P-ISSN: 2685-9653*.

Chen, Yifan et al. (2023). *Critical slip line recognition and extraction method of slope based on modified k-medoid clustering algorithm*: Computers and Geotechnics. https://doi.org/10.1016/j.compgeo.2022.105125

Narayana, Surya G. (2019). *Unsupervised Clustering Categorical Data Using Evolutionary Optimization Techniques*. UK : Lulu Press.

Nguyen, Vivian M et al. (2013). *Recreational anglers' attitudes, beliefs, and behaviors related to catch-and-release practices of Pacific salmon in British Columbia*: Journal of Environmental Management. https://doi.org/10.1016/j.jenvman.2013.06.010

Sureja, Nitesh et al. (2022). *An improved K-medoids clustering approach based on the crow search algorithm*: Journal of Computational Mathematics and Data Science. https://doi.org/10.1016/j.jcmds.2022.100034