

PERBANDINGAN METODE OPTIMASI UNTUK PENGELOMPOKAN PROVINSI BERDASARKAN SEKTOR PERIKANAN DI INDONESIA (Studi Kasus Dinas Kelautan dan Perikanan Indonesia)

Edy Sulistiyawan^{1*}, Alfisyahrina Hapsery², Lucky Junita Ayu Arifahanum³

¹ Program Studi Manajemen, Fakultas Ekonomi dan Bisnis, Universitas PGRI Adi Buana Surabaya

^{2,3} Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas PGRI Adi Buana Surabaya

*email: edy.sulistiyawan@gmail.com

ABSTRACT

The fisheries sector has an important role in supporting the food security chain, where the world's protein needs can be met by fisheries resources, both from capture fisheries and aquaculture. There are several fisheries sectors including fishing companies, capture fisheries production, number of ships, types and size of cultivated land. Therefore a statistical analysis is needed to increase the potential of fisheries in Indonesia. Data on the fisheries sector used in this study from the Indonesian Central Statistics Agency in 2018, which included the 2016 fisheries sector with 34 observation units in Indonesia. By using cluster analysis K-Means aims to group provinces in Indonesia based on the fisheries sector so that several groups are formed which will show the characteristics of each group. There are three determinations of the optimum number of clusters, namely the Elbow method, Silhouette method, and GAP Statistics. The results showed that optimum clusters were formed in 2 clusters, with the best Elbow and Silhouette methods. Where the first cluster is a region that shows a low value of the fisheries sector consisting of 30 provinces this is due to inadequate infrastructure and use that is not optimal while cluster 2 regions that have great potential in the Indonesian fisheries sector in 2016 as many as 4 provinces namely West Java, Java Central, East Java, and South Sulawesi as dominating capture fisheries production and aquaculture.

Keywords: Fisheries Sector, K-Means Cluster Analysis, Elbow Method, Silhouette Method and GAP Statistics.

1. PENDAHULUAN

Negara kepulauan yang memiliki berbagai potensi laut dan dikenal beranekaragam yaitu Indonesia. Potensi tersebut tersebar sepanjang kurang lebih 5,8 juta km zona maritim. Pembangunan ekonomi di Indonesia sangat berpotensi dengan didukung perikanan, terutama peluang lapangan pekerjaan. Kelautan sangat berperan penting untuk pemerataan pendapatan serta peningkatan taraf hidup nelayan kecil, dan para pengusaha dibidang perikanan (Kementerian Kelautan dan Perikanan, 2013).

Meningkatnya produksi hasil perikanan di Indonesia berkaitan dengan kondisi dan letak suatu daerah. Sebagaimana setiap wilayah memiliki kawasan penangkapan dengan karakteristik yang berbeda-beda. Untuk mengetahui karakteristik suatu kelompok pada sektor perikanan berdasarkan data masing-masing daerah dapat dianalisis menggunakan Ilmu statistika. Analisa multivariat yang dapat memahami, menjelaskan dan mengelompokkan seperangkat variabel adalah analisis *cluster*. Pada analisis ini, pengelompokan objek terjadi berdasarkan kemiripan antar objek (Supranto, 2004). Pengukuran kesamaan dari seluruh objek dapat diukur dengan ukuran kedekatan antar objek yang terdapat kedekatan ataupun tidak. Terdapat tiga metode dalam menentukan jumlah *cluster* yang optimum yaitu metode *Elbow*, metode *Silhouette*, dan *GAP Statistic*. Penentuan *cluster* optimum pada metode *Elbow* dapat dilihat dari nilai *Sum Square Error* (SSE), dimana jika nilai SSE semakin kecil maka nilai cluster tersebut terbaik. Berbeda dengan metode *Silhouette*, nilai koefisien *Silhouette* biasanya di antara rentang -1 sampai dengan 1. Nilai koefisien *Silhouette* yang mendekati 1 adalah yang lebih baik. Sedangkan *GAP Statistic* dikatakan *cluster* optimum jika nilai *GAP Statistic* mengalami kenaikan yang paling tinggi.

Pada penelitian sebelumnya yang berkaitan dengan analisis *cluster* adalah penelitian Prayoga & Ismaini Zain (2015) yang berjudul “Analisis faktor dan pengelompokan kecamatan berdasarkan indikator mutu pendidikan jenjang pendidikan dasar di kabupaten Sidoarjo”, terbentuk 3 kelompok pada jenjang SD dan 2 kelompok pada jenjang SMP dari 18 kecamatan Sidoarjo. Dengan adanya pemaparan diatas sehingga digunakan Analisis *Cluster Non Hirarki* untuk dapat pengelompokan serta membuat visualisasi pemetaan provinsi di Indonesia berdasarkan sektor perikanan sehingga terbentuk beberapa kelompok yang akan menunjukkan karakteristik anggota-anggota *cluster* yang terbentuk.

Berdasarkan penjelasan tersebut akan dilakukan pengelompokan sektor perikanan di Indonesia berdasarkan faktor-faktor yang mempengaruhinya antara lain jumlah rumah tangga/perusahaan tangkap, produksi perikanan tangkap, jumlah perahu/kapal, jumlah rumah tangga perikanan budidaya, jenis budidaya perikanan, produksi perikanan budidaya, luas area tambak usaha budidaya perikanan tambak, dan luas area kolam usaha budidaya perikanan. Dengan menggunakan analisis *cluster* untuk mengetahui karakteristik suatu anggota *cluster* yang terbentuk. Serta mengetahui metode mana yang terbaik dalam menentukan jumlah *cluster* pada sektor perikanan Indonesia. Hal ini yang melatarbelakangi penulis mengangkat sebuah penelitian skripsi yang berjudul “Perbandingan Metode Optimasi untuk Pengelompokan Provinsi Berdasarkan Sektor Perikanan di Indonesia”.

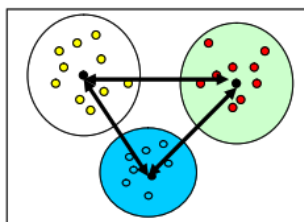
2. TINJAUAN PUSTAKA

2.1. Analisis dan Asumsi Cluster

Analisis *multivariate* yang mengelompokkan objek dikenal dengan sebutan *cluster*. Konsep dari metode ini adalah mengelompokkan objek kedalam kelompok berdasarkan homogenitas dari objek tersebut (Supranto,2004). Mengukur kedekatan diantara banyak objek dilakukan dengan perhitungan jarak menggunakan *Euclidean* (Johnson & Wichern, 1982). Analisis *cluster* dibagi menjadi dua metode yaitu metode hirarki dan metode *non hirarki*. Analisis *cluster* dengan menggunakan metode *non hirarki* adalah metode pengelompokan dengan menentukan banyak kelas terlebih dahulu. *K-Means* merupakan salah satu metode data *clustering non hierarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok.

2.2. K-Means Cluster

Perpartisian suatu objek pada suatu k kelompok dapat dilakukan dengan analisis *K-Means*. Setiap objek pengamatan akan mengelompok berdasarkan nilai rata-rata terdekat, dengan bantuan iterasi sesuai dengan algoritma untuk mendapatkan pusat kelompok (Prasetyo, 2012).



Gambar 2.1 Cluster K-Means

K-Means salah satu metode pengelompokan data non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data untuk meminimalisir varians kedalam kelompok sehingga data berkarakteristik sama di masukkan kedalam kelompok yang lain (Agusta, 2007).

2.3. Optimasi Jumlah Cluster

Bagian ini hanya berisikan dasar teori yang sangat dibutuhkan, yang sangat mendukung metode penelitian dan pembahasan. Teori yang bersifat umum tidak perlu dituliskan

a. Metode Elbow

Bagian ini hanya berisikan dasar teori yang sangat dibutuhkan, yang sangat mendukung metode penelitian dan pembahasan. Teori yang Metode *Elbow* merupakan metode untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik. Jika nilai *cluster* pertama dengan *cluster* kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai *cluster* tersebut terbaik. Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*. Berikut rumus SSE pada *K-means*.

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|_2^2$$

Dimana k adalah *Cluster*, X_i adalah data ke- i dalam *cluster* x , C_k adalah nilai rata-rata k *cluster*.

b. Metode Silhouette

Silhouette Coefficient (SC) merupakan metode yang digunakan untuk memvalidasi pengelompokan. *Silhouette* gabungan dari *cohesion* dan *separation* (Tan et al. 2006). *Cohesion* dan *separation* digunakan untuk mengukur akurasi pengelompokan berdasarkan kesamaan atau ketidaksamaan antar *cluster*. *Cohesion* digunakan untuk mengukur kedekatan data yang berada pada satu *cluster* sedangkan *separation* digunakan untuk mengukur kedekatan antar *cluster* yang terbentuk (Ramadhan, 2010). Nilai koefisien *Silhouette* biasanya di antara rentang -1 sampai dengan 1. Nilai koefisien *Silhouette* yang mendekati 1 adalah yang lebih baik. Hitung rata-rata jarak dari suatu objek misalkan i dengan semua objek lain yang berada dalam satu *cluster*.

$$a(i) = \frac{1}{|A|} \sum_{j \in A, j \neq i} Cd(i, j)$$

j adalah Objek lain dalam satu *cluster*, sedangkan $d(i, j)$ adalah jarak antar objek i dengan j . Hitung rata-rata dari objek i tersebut dengan semua objek di *cluster* lain, dan diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} Cd(i, j)$$

$d(i, C)$ adalah jarak rata-rata objek i dengan semua objek pada *cluster* lain atau C dimana $A \neq C$, sedangkan $b(i)$ adalah $\min_{C \neq A} d(i, C)$. Nilai koefisien *Silhouette* nya dihitung menggunakan persamaan ini.

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

c. GAP Statistic

GAP Statistic digunakan ketika akan membentuk *cluster* agar lebih hasil dapat konstan jika dibandingkan lainnya. Jarak objek berpasangan dalam *cluster* dihitung dengan rumus:

$$D_r = \sum_{i, i' \in C_r} dii'$$

d merupakan jarak *euclidean* kuadrat. Untuk menghitung jumlah kuadrat pada *cluster* menggunakan rumus

$$W_k = \sum_{r=1}^k \frac{1}{2n_r} D_r$$

nilai GAP didapatkan dengan mengestimasi jumlah *cluster* optimum pendekatan standardisasi W_k

$$Gap_n(k) = E_n^* \{ \log(W_k) \} - \log(W_k)$$

dimana E_n^* adalah nilai ekspektasi dari distribusi jumlah sampel. Kriteria banyak *cluster* optimum diberikan oleh nilai *GAP Statistic* (k) yang paling tinggi, atau yang pertama kali mengindikasikan kenaikan GAP yang minimum jika gap selalu naik (Tibshirani, dkk. 2001).

3. METODE PENELITIAN

3.1. Sumber Data dan Variabel

Data yang digunakan pada penelitian ini diperoleh dari Badan Pusat Statistik Indonesia tahun 2018. Data yang digunakan adalah data sektor perikanan di Indonesia tahun 2016 yang terdiri dari 34 provinsi. Penelitian ini menggunakan 10 variabel sektor perikanan Indonesia tahun 2016 seperti pada Tabel 1 berikut.

Tabel 1. Variabel Penelitian Sektor Perikanan Indonesia Tahun 2016

	Variabel Penelitian	Definisi Operasional
X ₁	Jumlah Rumah Tangga/Perusahaan Tangkap	Jumlah/total rumah tangga yang melakukan penangkapan ikan/binatang air lainnya/ tanaman air.
X ₂	Produksi Perikanan Tangkap	Hasil penangkapan ikan laut yang ada di provinsi (ton).
X ₃	Jumlah Perahu/Kapal	Jumlah kapal motor yang ada di provinsi.
X ₄	Jumlah Rumah Tangga Perikanan Budidaya tambak	Jumlah/total rumah tangga yang melakukan kegiatan budidaya ikan/binatang air lainnya/ tanaman air dengan media tambak.
X ₅	Jumlah Rumah Tangga Perikanan Budidaya kolam	Jumlah/total rumah tangga yang melakukan kegiatan budidaya ikan/binatang air lainnya/ tanaman air dengan media kolam.
X ₆	Jenis Budidaya Perikanan	Luas lahan jenis budidaya yang digunakan dengan satuan ha
X ₇	Produksi Perikanan Budidaya Tambak	Hasil penangkapan budidaya ikan tambak yang ada di provinsi (ton).
X ₈	Produksi Perikanan Budidaya Kolam	Hasil penangkapan budidaya ikan kolam yang ada di provinsi (ton).
X ₉	Luas Area Tambak Usaha Budidaya Perikanan Tambak	Luas lahan tambak yang digunakan dalam kegiatan budidaya dengan satuan ha.
X ₁₀	Luas Area Kolam Usaha Budidaya Perikanan	Luas lahan kolam yang digunakan dalam kegiatan budidaya dengan satuan ha.

3.2. Metode Analisis Data

Langkah-langkah analisis pada penelitian ini dapat dipaparkan sebagai berikut :

1. Melakukan analisis statistika deskriptif untuk mengkaji karakteristik sektor perikanan pada tahun 2016 dengan menggunakan penyajian tabel.
2. Melakukan standarisasi data dengan menggunakan *Z-Score*.
3. Menentukan *cluster* optimum terbaik dengan menggunakan metode *Elbow*, metode *Silhouette*, dan *GAP statistic*.
4. Menentukan *centroid* dari data pada setiap *cluster*
5. Menentukan jarak pada setiap objek dengan setiap *centroid* untuk menghitung jarak disetiap objek dengan *centroid* berdasarkan *Euclidean Distance*
6. Mengklasifikasikan setiap data berdasarkan kedekatan dengan *centroid* (jarak minimum)

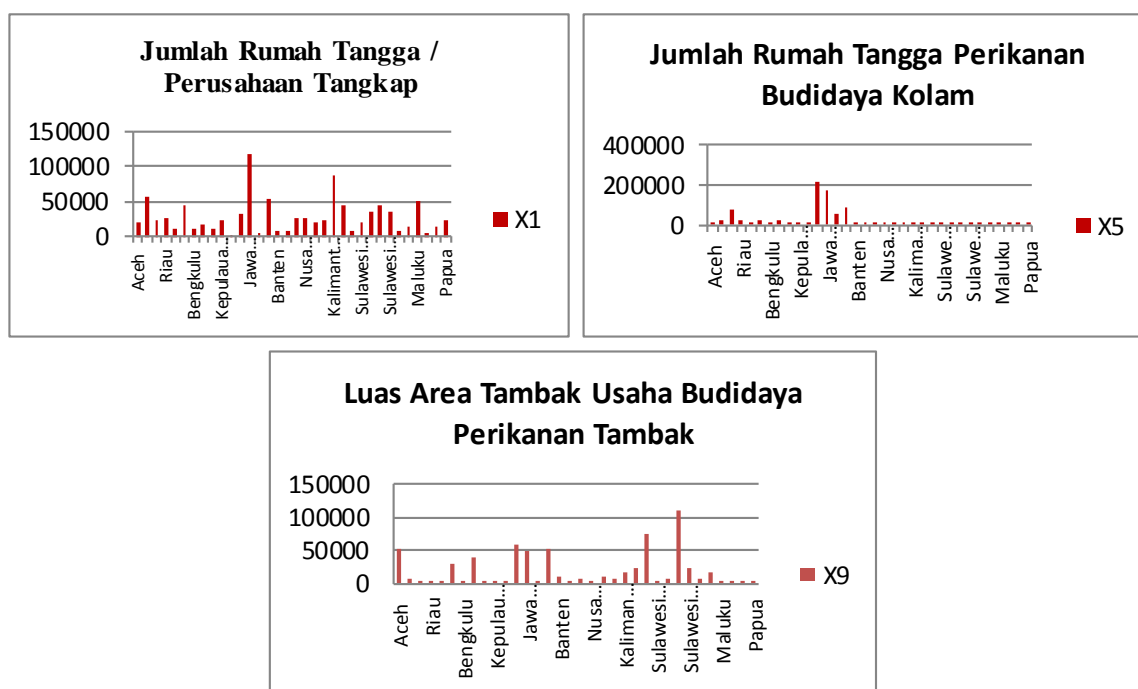
7. Membagi wilayah sesuai dengan jumlah *cluster* yang terbentuk
8. Mengintepretasikan karakteristik anggota *cluster* yang terbentuk
9. Kesimpulan

4. HASIL DAN PEMBAHASAN

Analisis dan pembahasan pada sektor perikanan Indonesia diawali dengan melihat karakteristik data sektor perikanan, Pengujian asumsi analisis cluster, Pengelompokan wilayah berdasarkan K-means Cluster, dan Visualisasi data dari hasil pengelompokan wilayah berdasarkan sektor perikanan di Indonesia.

4.1. Karakteristik Sektor Perikanan

Karakteristik sektor perikanan Indonesia pada tahun 2016 dapat dilihat pada diagram batang pada Gambar 1 berikut.



Gambar 1. Diagram Batang (a) Jumlah Rumah Perusahaan Tangkap, (b) Jumlah RT Perikanan Budidaya Kolum, (c) Luas Area Tambak Usaha Budidaya Perikanan Tambak.

Gambar 1 menjelaskan karakteristik dari beberapa variabel. Jumlah rumah tangga (RT)/perusahaan tangkap di Indonesia pada provinsi Jawa Tengah mencapai 118.939 RT yang merupakan jumlah tertinggi di Indonesia pada tahun 2016. Rata-rata jumlah rumah tangga (RT)/perusahaan tangkap sebesar 28.413. Jumlah rumah tangga perikanan budidaya kolam yang tertinggi di Indonesia pada tahun 2016 terjadi pada provinsi Jawa Barat sebesar 218.335 rumah tangga (RT). Sedangkan untuk jumlah rumah tangga perikanan budidaya kolam terendah terjadi pada provinsi Maluku mencapai 179 RT. Indonesia memiliki rata-rata untuk jumlah rumah tangga perikanan budidaya tambak sebesar 26.215 RT. Luas area tambak usaha budidaya perikanan di Indonesia pada tahun 2016 memiliki luas terbesar pada provinsi Sulawesi Selatan mencapai 111.039 ha. Sedangkan Provinsi Sumatra Barat memiliki luas area tambak usaha budidaya perikanan terendah mencapai 17 ha. Rata-rata luas area tambak usaha budidaya perikanan di Indonesia pada tahun 2016 mencapai 17.821 ha.

4.2. Pengujian Asumsi Analisis Cluster

Beberapa tahapan sebelum melakukan analisis *cluster K-Means* yaitu uji asumsi dan penentuan jumlah cluster optimum uji asumsi kecukupan data dan multikolinieritas sebagai berikut.

a. Uji Asumsi Kecukupan Data

Pengujian kecukupan data dapat dilihat dari hasil pengujian dengan *Kaiser Mayer Olkin* (KMO). Dimana pendugaan awal adalah jumlah sampel tidak memenuhi kriteria untuk analisa lebih lanjut.

Tabel 3 Uji *Kaiser Mayer Olkin*

Keterangan	Nilai
KMO Measure of Sampling Adequacy	0.72
Bartlett's Test	289.33

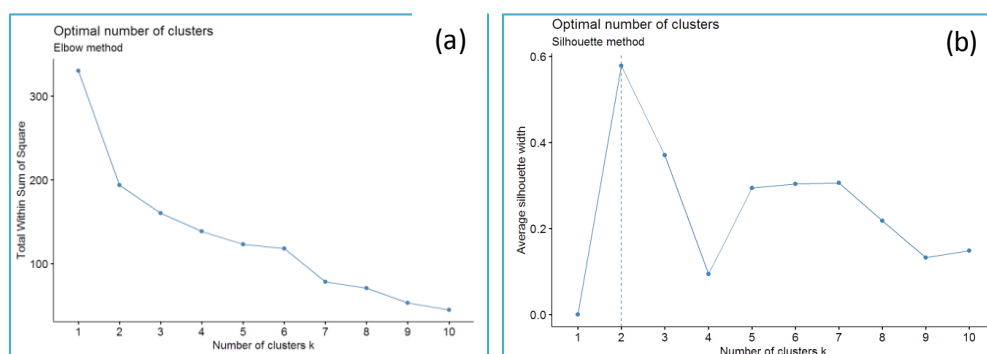
Tabel 3 memberikan informasi nilai KMO sebesar 0,72. Hal tersebut menunjukkan bahwa nilai 0.72 lebih besar dari 0.5, yang berarti tolak H_0 sehingga kesimpulan yang diambil adalah sampel telah memenuhi untuk dilakukan analisis.

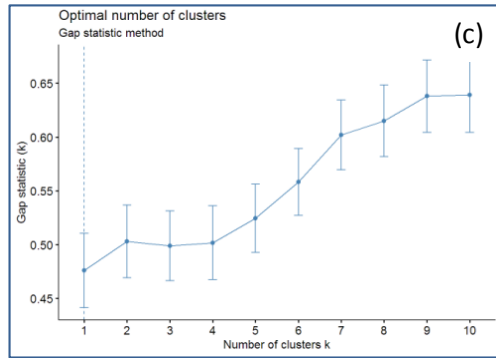
b. Uji Multikolinieritas

multikolinieritas dapat diketahui dari nilai koefisien pada pengujian korelasi sederhana (*Correlation Pearson*). Hasil analisis menunjukkan adanya korelasi antar variabel *independent* yang tertinggi sebesar 0.898 yang didapatkan dari hasil korelasi antara variabel produksi perikanan budidaya kolam (X_8) dengan luas area kolam usaha budidaya perikanan (X_{10}). Karena nilai 0.898 kurang dari 0.9 maka dinyatakan bahwa data penelitian tidak terdeteksi adanya multikolinieritas.

4.3. Perbandingan Metode Optimasi pada Analisis Cluster

Pada metode *K-Means* terdapat beberapa pertimbangan untuk menentukan jumlah *cluster* yang paling optimal. Pendekatan metode yang banyak digunakan adalah Metode *Elbow*, Metode *Silhouette*, dan *GAP Statistic*. Hasil optimasi dari ketiga metode untuk Elbow dapat dilihat pada Gambar 4 berikut.





Gambar 2. (a) Grafik Method *Elbow*, (b) Grafik Metode *Silhouette* (c) Grafik Metode *GAP Statistic*

Penentuan jumlah kelompok optimal dari metode elbow diperoleh dari nilai *within Sum of Square*. Grafik metode *Elbow* menunjukkan penurunan yang signifikan dan membentuk siku pada titik $k=2$. Sehingga dapat disimpulkan pada metode ini jumlah cluster optimal adalah 2. Metode kedua adalah metode *Silhouette* dapat dilihat pada Gambar 4 (b). Penentuan k optimal pada metode *Silhouette* dapat diketahui dengan melihat nilai *average Silhouette*, dengan asumsi semakin tinggi nilai *average* maka k tersebut semakin optimal. Pada gambar 4(c) tersebut menunjukkan bahwa nilai *average Silhouette* tertinggi berada pada $k=2$. Maka, dapat disimpulkan bahwa dengan menggunakan metode *Silhouette* diperoleh k optimalnya yaitu pada $k=2$. Metode terakhir untuk optimasi adalah metode *GAP Statistic*. Pada gambar 4 grafik *GAP Statistic* menunjukkan titik k berada pada $k=1$ yang merupakan jumlah optimal untuk membentuk *cluster*.

4.4. Identifikasi Cluster dan Visualisasi Data

Penentuan nama *cluster* dapat dilihat berdasarkan perbandingan nilai *mean* masing-masing variabel untuk setiap cluster. Pada penelitian ini, rata-rata sektor perikanan Indonesia menurut *cluster* seperti pada Tabel 4.

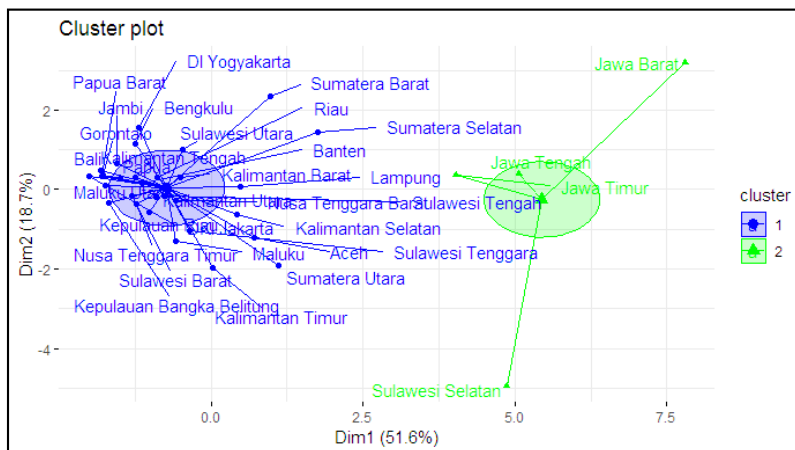
Tabel 4. Rata-Rata Sektor Perikanan Indonesia Tahun 2016

Variabel	Cluster	
	1	2
X ₁	23925.11	62072.13
X ₂	175922.04	325632.22
X ₃	4650.71	8053.21
X ₄	4088.78	35217.62
X ₅	13261.87	123363.46
X ₆	20271.32	148287.09
X ₇	29342.41	533048.85
X ₈	42755.07	251574.00
X ₉	11263.25	67004.15
X ₁₀	3782.84	13533.72

Tabel 4 diatas dapat dilakukan profilisasi untuk setiap kolom yang terbentuk. Karakteristik pada *cluster* 1 yang memiliki anggota 30 provinsi memiliki nilai rata-rata lebih rendah dari *cluster* 2. Sedangkan pada *cluster* 2 yang beranggota 4 provinsi merupakan daerah yang

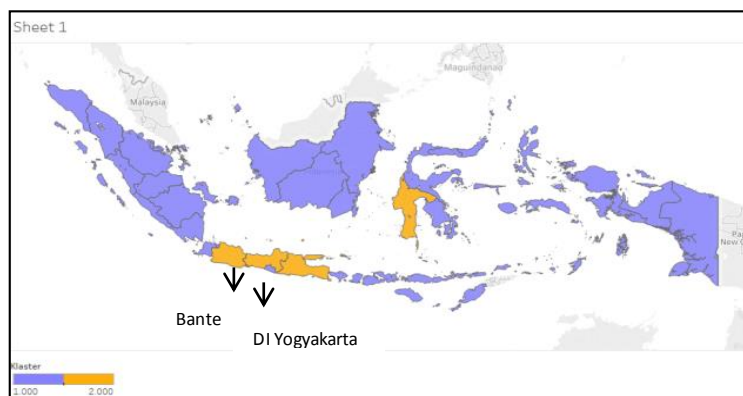
memiliki potensi sektor perikanan yang tinggi terdapat di provinsi Jawa Timur, Jawa Tengah dan Jawa Barat.

Untuk mengetahui anggota setiap *cluster* digunakan selisih jarak antar objek terhadap masing-masing *centroid*. Nilai dari *centroid* pertama dapat dipilih secara acak perhitungan jarak dilakukan dengan menggunakan persamaan jarak *Euclidean* pada persamaan 2, dimana objek dengan jarak terdekat dijadikan satu gerombol.



Gambar 5. Plot Pengelompokan Provinsi Indonesia

Gambar 5 diketahui bahwa dengan jumlah *cluster* optimal 2, terbentuk *cluster* 1 yang terdiri dari 30 provinsi, sedangkan *Cluster* 2 terdiri dari 4 provinsi. Pada bagian barat pulau Jawa terdapat suatu provinsi yaitu Banten, namun hal ini tidak menunjukkan adanya pencapaian yang positif pada sektor perikanan. Pada dasarnya tidak adanya kapasitas layanan seperti pelabuhan ikan yang sesuai untuk potensi wilayah yang ada menjadi salah satu alasan tidak tercapainya pemanfaatan zona penangkapan ikan (ZEEI).



Gambar 6. Pengelompokan Provinsi Berdasarkan Analisis *Cluster*

Selain Banten, juga terdapat provinsi DI Yogyakarta yang masih tergolong rendah akan sektor perikanan, dilihat dari penduduknya sebagian besar bermata pencarian sebagai petani. Apabila ditinjau secara luas, dapat dikatakan tradisi menangkap ikan bagi masyarakat, dengan rata-rata trip untuk melaut per bulan hanya 20 hari. Begitu pula dengan pemabangunan sarana pelabuhan masih kurang mendukung.

5. KESIMPULAN

Indonesia memiliki 34 provinsi, berdasarkan data sektor perikanan di Jawa Timur memenuhi 9 variabel sektor perikanan. Hasil pengujian *Kaiser Mayer Olkin* menunjukkan

nilai signifikansi 0.72 yang berarti sampel yang digunakan sudah cukup untuk dianalisis lebih lanjut dan memiliki nilai Korelasi *Pearson* tertinggi sebesar 0.898 kurang dari 0.9 menyatakan bebas multikolinearitas. Pada perbandingan optimasidari ketiga metode, diketahui cluster optimal dalam pembentukan cluster sebanyak 2 cluster dengan metode *Elbow* dan metode *Silhouette*.

Cluster pertama terdiri dari 30 provinsi, di antaranya Provinsi DI Yogyakarta, Banten, dan seluruh provinsi yang ada di pulau Sumatra, Pulau Kalimantan, Pulau Sulawesi, Nusa Tenggara, Maluku, serta Papua. Dimana cluster pertama ini tergolong sektor perikanan yang masih rendah, hal ini disebabkan infrastruktur masih belum memadai, pemanfaatan yang belum optimal, serta potensi sumber daya alam lainnya yang lebih memberikan kontribusi besar terhadap pendapatan daerah. Sedangkan *cluster* kedua terdiri dari 4 provinsi yaitu Jawa Timur, Tengah, dan Barat yang terletak di pulau Jawa sehingga secara geografis berhadapan dengan dua sisi lautan jawa pada bagian utara, dan samudra Hindia sehingga mendatangkan keuntungan yang besar dari produksi perikanan tangkap hasil laut.

DAFTAR PUSTAKA

- Ediyanto, Mara M.N., Satyahadewi N., (2013), Pengklasifikasian Karakteristik Dengan Metode *K-Means Cluster Analysis*, *Bulletin Ilmiah Matematika Statistika dan Terapannya*, 2(2): 133-136.
- FAO *Fisheries Departement*, (2006), *State of world aquaculture 2006*, FAO Fisheries Technical Paper, No.500. Rome, 134p.
- Kementerian PPN, (2014), *Kajian Strategi Pengelolaan Perikanan Berkelanjutan*, Jakarta: Direktorat Kelautan dan Perikanan 2014.
- Prayoga, A. & Zain, I. (2015). *Analisis Faktor dan Pengelompokan Kecamatan berdasarkan Indikator Mutu Pendidikan Jenjang Pendidikan Dasar di Kabupaten Sidoarjo*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Supranto, J., (2004), *Analisis Multivariat Arti dan Interpretasi*, Jakarta: Rineka Cipta.
- Yana, Mira Suci, dkk, (2018), *Penerapan metode K-Means dalam pengelompokan wilayah menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018*. Banda Aceh: Jurusan Statistika, FMIPA, Universitas Syiah Kuala.