

CLUSTERING PRODUKSI PERIKANAN BUDIDAYA LAUT BERDASARKAN PROVINSI MENGGUNAKAN ALGORITMA K-MEANS

Alfian Firlansyah¹, Andi Akram Nur Risal², Fhatiah Adiba³, Andi Baso Kaswar⁴

alfianfirlansyah1829@student.unm.ac.id, akramandi@unm.ac.id, fhatiahadiba@unm.ac.id,

a.baso.kaswar@unm.ac.id

^{1,2,3,4} Universitas Negeri Makassar

Received: 28 Apr 2021
Accepted: 14 May 2021
Published: 15 May 2021

Abstract

Abstract: Indonesia is a large maritime country, and most of its territorial waters are larger than its land area. Due to the vastness of the oceans, the large number of large and small islands makes Indonesia a potential area for marine cultivation. In general, the existing data based on the Central Statistics Agency (BPS) of Marine Aquaculture Production for each province in Indonesia only applies to production data which only produces detailed data on total marine aquaculture production in tonnes per year, and takes a long time. To classify very large data, a method is needed that can use the K-Means algorithm to classify the highest, middle, and lowest opportunities in the field of marine aquaculture from 2004 to 2018. The results implemented in python consisted of 26 provinces in klaster 1 (C1), 3 provinces in klaster 2 (C2), and 5 provinces in klaster 3 (C3).

Keywords: Indonesia, Classify, K-Means, Python, Provinces

Abstrak

Indonesia merupakan negara maritim yang besar, dan sebagian besar wilayah perairannya lebih besar dari luas daratan. Karena luasnya lautan, banyaknya pulau besar dan kecil menjadikan Indonesia sebagai daerah yang potensial untuk budidaya laut. Secara umum, data yang ada berdasarkan Badan Pusat Statistik (BPS) Produksi Akuakultur Laut masing-masing provinsi di Indonesia hanya berlaku untuk data produksi yang hanya menghasilkan data detail total produksi perikanan budidaya laut dalam ton per tahun, dan memakan waktu lama. Untuk mengelompokan data yang sangat besar, maka diperlukan suatu metode yang dapat menggunakan algoritma *K-Means* untuk mengklasifikasikan peluang tertinggi, menengah, dan terendah pada bidang budidaya perikanan laut dari tahun 2004 hingga 2018. Hasil yang diimplementasikan pada *python* terdiri dari 26 provinsi di klaster 1 (C1), 3 provinsi di klaster 2 (C2), dan 5 provinsi di klaster 3 (C3).

Kata kunci: Indonesia, Klasifikasi, K-Means, Python, Provinsi

This is an open access article under the
CC BY-SA license



1. Pendahuluan

Indonesia merupakan negara maritim yang besar, dan sebagian besar wilayah perairannya lebih besar dari luas daratan. Wilayah laut merupakan wilayah terluas dengan keanekaragaman biota laut yang memiliki keunikan berbeda-beda dengan wilayah daratan Indonesia lainnya.

Karena luasnya lautan, banyaknya pulau besar dan kecil menjadikan Indonesia sebagai daerah yang potensial untuk budidaya laut. Perkembangan budidaya laut disebut juga sebagai kelanjutan pengembangan budidaya laut. Dengan dukungan potensi yang cukup besar, budidaya laut juga dinilai mampu memberikan kontribusi yang besar untuk mendorong Indonesia menjadi pusat maritim dunia. Budidaya laut merupakan salah satu bidang yang diharapkan dapat mendukung peningkatan kesejahteraan masyarakat Indonesia. Karena besarnya potensi jumlah dan jenis sumber daya ikan, sub sektor perikanan budidaya laut dapat berperan dalam pemulihan dan pertumbuhan ekonomi.

Berkaitan dengan pentingnya perkembangan budidaya laut di Indonesia maka pengelolaan data juga sangat diperlukan. Oleh karena itu, dalam menentukan provinsi dengan peluang budidaya perikanan laut terbaik, pengelolaan data didasarkan pada data produksi rata-rata tahunan (ton).

Data mining merupakan metode yang digunakan dalam pengolahan informasi berskala besar oleh sebab itu informasi mining mempunyai peranan yang sangat berarti dalam sebagian bidang kehidupan antara lain ialah bidang industri, bidang keuangan, cuaca, ilmu serta teknologi [1].

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih kelompok [2]. Algoritma *K-Means* akan menguji setiap komponen dalam populasi data dan menandai komponen ini sebagai salah satu pusat cluster yang telah ditentukan, tergantung pada jarak minimum antara komponen dan setiap pusat cluster. Selain itu, lokasi pusat cluster akan dihitung ulang hingga semua komponen data diklasifikasikan ke dalam masing-masing cluster, dan akan terbentuk cluster baru di bagian akhir.

Sebelumnya telah dilakukan penelitian pengelompokan dibidang pariwisata dengan menggunakan algoritma yang serupa yaitu, *K-Means* [1][3][4][5].

Untuk memperoleh pengelompokan dari data kunjungan wisatawan hotel berbintang di berbagai provinsi dari tahun 2003 hingga 2016, dilakukan

pengelompokan perjalanan mancanegara ke Indonesia menurut provinsi [1]. Selain itu, penelitian klasifikasi wisata Kota Yogyakarta menggunakan algoritma *K-Means* untuk mengetahui obyek wisata mana yang memiliki potensi kunjungan paling rendah hingga volume kunjungan tertinggi. Dengan 5 iterasi menghasilkan cluster 1 berisi 24 data, cluster 2 ada 11 data, dan Cluster 3 memiliki 13 data [3]. Kemudian, pengelompokan jadwal perjalanan biro perjalanan dan biro perjalanan tersebut. Ada tiga jenis rute perjalanan, yaitu, rute perjalanan paling populer berjumlah 20%, rute perjalanan terpopuler mencapai 30%, dan rute paling tidak populer menyumbang 50% [4]. Kemudian mempelajari pengelompokan kunjungan pengunjung ke destinasi utama DKI Jakarta. Hasilnya dikelompokkan menjadi C1 = Taman Impian Jaya Ancol, C2 = Taman Mini Indonesia Indah dan Kebun Binatang Ragunan, dan C3 = Monumen Nasional, Museum Nasional, Museum Nasional, Museum Satria Mandala, Museum Sejarah Jakarta Museum dan Pelabuhan Sunda Kelapa. Hasil pengelompokan C3 menjadi catatan pemerintah provinsi DKI. Jakarta [5].

Adapun penelitian yang menggunakan algoritma serupa yang digunakan dalam segmentasi untuk mendeteksi citra penyakit daun tanaman jagung. proses segmentasi berbasis *K-Means* sebagai input dengan menentukan jumlah cluster awal adalah $k=3$, merandom centroid, menghitung jarak nilai pixel ke centroid, mengelompokkan nilai pixel berdasarkan jarak minimum, menghitung rata-rata cluster untuk centroid baru dan jika masih terdapat nilai pixel yang berpindah maka proses random centroid masih dilakukan hingga tidak adanya nilai pixel yang berpindah. Penelitian tersebut menggunakan dataset sejumlah 30 jenis A dan B untuk training dan 10 data untuk testing telah memperoleh akurasi sebesar 90% [6].

Selanjutnya, penelitian pengelompokan dibidang pendidikan dalam pemilihan jurusan di SMK swasta harapan baru. Setiap data siswa dihitung jarak kedekatannya dengan centroid dari masing-masing jurusan, setelah itu dilakukan update terhadap nilai centroid berdasarkan nilai rata-rata dari masing-masing kelompok. Bila nilai centroid masih berubah, maka dilakukan perhitungan jarak ulang hingga nilai centroid tidak berubah, dan data clustering stabil. Hasil penelitian adalah aplikasi dapat digunakan untuk membantu proses pemilihan jurusan pada siswa SMK Swasta Harapan Baru secara otomatis, sehingga dapat membantu siswa/siswi yang masih bingung dengan pemilihan jurusan dan dapat menghindari kesalahan dalam pemilihan jurusan [7].

Adapun penelitian pengelompokan yang serupa mengenai hasil tangkap ikan di pelabuhan perikanan nusantara (PPN) Ternate menggunakan algoritma *K-Means* untuk mengetahui seberapa bagus atau baik klaster yang dipakai pada penelitian, penulis sudah melakukan perhitungan beberapa klaster. Hasil penelitian menggunakan 18 data ikan dengan 2 klaster, dimana ada 16 data yang masuk pada klaster satu (C1) dan 2 data yang masuk pada klaster dua (C2) [8].

Penelitian mengenai pengelompokan dengan menggunakan modifikasi *spatio-temporal density based* algoritma *clustering* yaitu ST-DBSCAN. Penelitian pengelompokan tambahan pada data titik api terhadap hutan dan lahan sebagai indikator kebakaran di Sumatera menghasilkan pengelompokan awal pada parameter ϵ dari 0,1, ϵ dari 7, dan $MinPts$ dari 5 menghasilkan 16 cluster dan 23 *outlier*. *Inkremental clustering* pada parameter tersebut menghasilkan 7 klaster baru dan 6 *outlier* baru [9].

Dari beberapa penelitian yang telah dipaparkan sebelumnya, dapat disimpulkan bahwa dalam mengelompokkan data, algoritma yang sangat cocok digunakan ialah *K-Means* dengan menentukan nilai K yang terbaik untuk data tersebut sehingga menghasilkan pengelompokan data yang sesuai.

Pada penelitian ini, penulis mengusulkan Pengelompokan Produksi Perikanan Budidaya Laut Berdasarkan Provinsi Menggunakan Algoritma *K-Means* yang diimplementasikan pada *Python*.

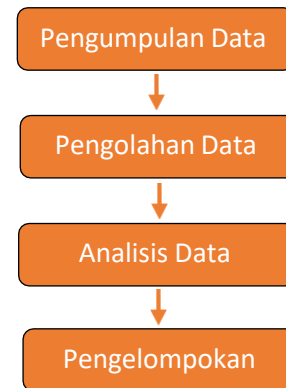
Pada umumnya data yang ada pada badan pusat statistik produksi perikanan budidaya laut berdasarkan provinsi di Indonesia hanya menerapkan data produksi yang hanya menghasilkan data detail total produksi perikanan budidaya laut dalam satuan ton disetiap tahunnya dan butuh waktu yang cukup lama dalam mengelompokkan data berukuran yang sangat besar sehingga, sulitnya dalam mengelompokkan provinsi dengan peluang tertinggi, menengah, dan terendah dibidang perikanan budidaya laut.

Tujuan dari penelitian ini adalah pengolahan data dengan metode yang dapat mengelompokkan peluang tertinggi, menengah, dan terendah dibidang perikanan budidaya laut 2004-2018 dengan menggunakan algoritma *K-Means*. Sehingga, hasil dari pengolahan data tersebut memberikan informasi kepada nelayan dan orang-orang yang bekerja dibidang perikanan tentang provinsi dengan peluang tertinggi untuk budidaya laut.

2. Metode

Metode dalam penelitian ini terdiri dari tahap pengumpulan data hingga tiga tahapan utama yaitu:

pengolahan data, analisis, dan pengelompokan. Pada tahap pengelompokan diterapkan algoritma *K-Means* sehingga dapat dikelompokkan berdasarkan klaster dengan peluang tertinggi (C1), terendah (C2), dan menengah (C3) berdasarkan Gambar 1.



Gambar 1. Metode Penelitian

2.1. Tahap Pengumpulan Data

Pengumpulan data produksi perikanan budidaya laut berdasarkan provinsi diperoleh dari situs Badan Pusat Statistik (BPS) dengan alamat url <https://www.bps.go.id/statictable/2009/10/05/1706/pr-duksi-perikanan-budidaya-menurut-provinsi-dan-jenis-budidaya-2000-2018.html>. Data yang digunakan dalam penelitian ini adalah data Produksi Perikanan Budidaya Menurut Provinsi dan Jenis Budidaya (ton), 2000-2018*. Data tersebut berisikan total produksi perikanan pada masing-masing provinsi yang ada di Indonesia dengan jenis-jenis budidaya dalam satuan ton Tahun 2000-2018. Pada penelitian ini dilakukan dengan menggunakan algoritma pengelompokan yaitu metode *K-Means*.

2.2. Tahap Pengolahan Data

Data yang telah diperoleh akan diolah terlebih dahulu untuk dapat dipengelompokkan. Dikarenakan data yang akan dikelompokkan adalah data Produksi Perikanan Jenis Budidaya Laut 2004-2018*, maka dilakukan pengolahan data dengan mengambil data yang dianggap penting saja untuk di pengelompokan. Sehingga, diperoleh data pada Tabel 1 yang akan diproses pada tahap analisis.

2.3. Tahap Analisis

Pada tahapan ini dilakukan analisis data Produksi Perikanan Budidaya Menurut Provinsi Jenis Budidaya Laut (ton), 2004-2018* seperti pada Tabel 1. Data yang diperoleh akan dianalisis dengan

menentukan jumlah pengelompokan yang sesuai untuk ke tahap pengelompokan. Dengan menganalisis telah ditentukan data tersebut akan dikluster ke dalam

3 peluang kluster yaitu, tertinggi, menengah, dan terendah.

Tabel 1. Presentase Produksi Perikanan Budidaya Menurut Provinsi dan Jenis Budidaya (Ton), 2000-2018*

Provinsi	Budidaya Laut																	
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	
Sumatera Utara	496	548	-	612	352	1.086	1.088	1.056	3.086	4.348	4.363	2.800	7	138	-	-	-	
Sumatera Barat	120	136	40	34	34	60	13	79	833	335	243	268	320	0	-	-	-	
Nusa Tenggara Barat	3.048	26	99	5	4	11	2	2	5	493	419	489	13.612	8.524	-	-	-	
Jambi	-	-	-	-	-	-	-	-	-	421	23	-	-	-	-	-	-	
Sumatera Selatan	-	-	-	-	346	357	392	-	-	-	-	-	-	-	-	-	-	
Bengkulu	-	-	-	-	800	-	-	248	5.844	770	24	1	64	1	120	-	-	
Lampung	1.099	821	1.099	2.094	1.473	4.201	8.448	10.006	14.057	13.927	4.711	4.440	3.889	4.848	2.842	-	-	
Kepulauan Bangka Belitung	31	24	31	24	42	732	786	4.381	10.142	7.342	1.084	1.064	1.013	811	170	-	-	
Kepulauan Riau	4.816	909	8.805	6.624	6.616	16.477	6.682	-	-	-	16.368	10.049	10.049	10.339	3.004	-	-	
DKI Jakarta	1.093	1.809	1.457	1.345	1.530	1.534	10.281	24.724	3.071	2.517	4.064	2.464	1.034	1.200	194	-	-	
Jawa Barat	10.000	10.000	10.007	10.070	10.035	8.433	14.707	7.844	3.736	3.528	1.011	3.064	8.519	400	4.894	-	-	
Jawa Tengah	-	20.984	21.512	1.814	1.244	1.408	1.408	1.573	6.004	16.211	10.413	17.785	1.812	18.142	10.414	-	-	
DI Yogyakarta	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Jawa Timur	219	4.012	10.148	13.011	71.718	130.487	388.015	812.790	541.097	583.018	621.411	415.462	410.319	110.121	171.769	-	-	
Bali	2.917	3.940	6.027	61.200	10.944	1.822	1.034	16.708	17.219	21.839	15.073	16.110	10.411	11.023	-	-	-	
Ban	150.054	161.121	164.709	152.306	128.174	130.000	90.880	106.647	146.610	146.140	88.914	107.921	101.706	104	1.064	-	-	
Nusa Tenggara Barat	30.048	34.212	40.041	70.516	86.232	147.046	163.287	278.107	451.892	549.742	749.624	921.549	1.001.291	823.749	820.242	-	-	
Nusa Tenggara Timur	442.488	271.880	481.121	504.700	499.278	499.428	347.924	277.000	289.708	1.846.262	1.306.262	2.241.247	1.810.461	1.841.704	1.840.860	-	-	
Kalimantan Barat	18	5.047	312	86	107	82	197	20	147	177	1.631	32	34	20	7	-	-	
Kalimantan Tengah	-	-	-	34	34	31	300	149	230	88	159	364	364	0	-	-	-	
Kalimantan Selatan	496	105	2.823	6.065	1.742	1.811	1.802	2.801	2.716	2.424	2.704	488	475	424	707	-	-	
Kalimantan Timur	56	81	1.743	14.444	4.006	7.587	15.995	83.214	195.510	248.748	321.124	22.124	1.713	27.543	13.115	-	-	
Kalimantan Utara	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Sulawesi Utara	3.705	3.739	4.630	4.500	4.827	4.847	48.144	101.544	151.642	174.052	301.876	339.124	201.149	119.411	200.611	-	-	
Sulawesi Tengah	13.700	134.112	170.275	182.074	284.204	710.991	716.494	716.494	891.428	1.214.021	1.137.084	1.274.904	1.240.041	923.141	1.144.274	-	-	
Sulawesi Selatan	201.41	201.410	486.474	415.727	461.349	621.349	815.727	1.024.310	1.480.791	1.841.417	2.087.474	2.411.814	2.247.447	2.627.200	2.241.211	-	-	
Sulawesi Tenggara	84.514	121.110	240.276	412.127	124.816	184.144	312.411	384.744	440.144	540.144	640.144	640.144	640.144	640.144	640.144	-	-	
Gorontalo	1.233	1.514	4.122	7.117	11.176	48.243	44.077	80.140	91.482	144.047	24.911	17.740	10.340	10.400	14.041	-	-	
Sulawesi Barat	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	1.149	-	-	
Makassar	3.992	302	3.912	171.911	27.066	52.134	271.911	449.712	476.176	586.106	490.091	713.409	391.149	491.149	644.449	-	-	
Makau Utara	124	834	714	1.080	1.480	2.288	48.978	45.091	121.110	86.312	103.213	41.502	240.029	714.51	146.622	-	-	
Papua Barat	18	1.647	77	39	204	149	174	176	355.556	146	11	101	112	400	144	-	-	
Papua	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Indonesia 400.019 890.074 1.361.018 1.009.120 1.066.002 1.202.083 1.514.702 4.045.827 4.792.717 6.274.211 6.084.716 10.174.022 9.773.015 8.084.014 1.341.515
 Catatan: * Angka dibulatkan
 Sumber: Direktorat Jenderal Perikanan Budidaya
 * Angka Sumatera/Provinsi/Regensi

Setelah dianalisis, Kemudian data tersebut yang akan digunakan untuk ke tahap pengelompokan dengan menggunakan algoritma *K-Means* dengan mengelompokkan 3 kelas.

2.4. Pengelompokan

Pada tahapan ini dilakukan pengelompokan terhadap data yang telah dianalisis dengan menggunakan algoritma *K-Means*.

Algoritma *K-Means* mampu meminimalkan jarak antara data ke klasternya. Pada dasarnya penggunaan algoritma ini dalam proses pengelompokan tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses [10].

Sehingga pada *K-Means* sangat penting menentukan jumlah cluster dan atribut yang betipe numeric.

Pada dasarnya *K-Means* terdiri dari 2 proses, yaitu pendeteksian lokasi pusat kluster dan proses pencarian anggotanya.

Proses *K-means* dimulai dengan menentukan jumlah kluster dan nilainya menggunakan jarak Euclidean dengan menghitung setiap data ke pusat kluster pada persamaan (1).

$$d_{ik} = \sum^m (C_{ij} - C_{kj})^2 \quad (1)$$

Lalu, Mengelompokkan data ke dalam kluster dengan mencari jarak terdekat pada persamaan (2).

$$\min \sum^i -d_{ik} = \sum^m (C_{ij} - C_{kj})^2 \quad (2)$$

k j

Selanjutnya, Hitung pusat kluster terbaru pada persamaan (3).

$$\sum^p C_{kj} = \frac{\sum^i x_{ij}}{p} \quad (3)$$

Dengan x_{ij} merupakan kluster ke-K dan p merupakan banyaknya anggota pada kluster ke-K

Lalu lakukan perulangan langkah pada persamaan 1 hingga persamaan 3 hingga tidak ada lagi data yang pindah dari kluster sebelumnya

3. Hasil dan Pembahasan

Pada penelitian ini, dilakukan pengolahan data yang mengelompokkan 3 kelas peluang produksi perikanan budidaya laut dengan menggunakan algoritma *K-Means*. Pada umumnya algoritma ini

terdiri dari 2 tahap utama, yaitu Centroid data, dan pengelompokan data. Data tersebut juga akan diimplementasi pada *Python*.

3.1. Mengimpor Paket

Pada tahap ini mengimpor beberapa package pada *Python* dan di inialisasi agar memudahkan dalam memanggil fungsi pada package yang telah di import. Package yang diimport pada pengelompokan ini ialah, *pandas*, *numpy*, *seaborn*, *matplotlib*, *sklearn*, *KMeans* dan *MinMaxScaler*.

3.2. Membaca Data

Setelah mengimpor paket, data yang akan di kluster dibaca terlebih dahulu dengan fungsi *read* sehingga dapat mengolah data pada tahap selanjutnya.

3.3. Mengolah Data

Pada tahap ini membaca data yang dianggap penting saja yang dimana berisikan data *numeric*. Sehingga mengarahkan program dengan fungsi *iloc* dalam membaca citra pada atribut 2004 – 2018. Lalu memberikan jarak antara data satu dengan data yang lain dengan fungsi *MinMaxScaler* untuk meminimalkan jarak setiap data yang terlalu jauh.

3.4. Mengelompokkan Data

Tahap selanjutnya ialah dengan menggunakan fungsi *K-Means* dengan menentukan jumlah kluster =

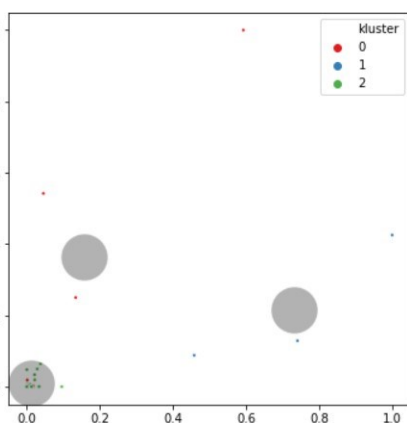
3 yang dimana nilai pada masing-masing kluster merupakan centroid. Sehingga menghasilkan pengelompokan seperti pada Tabel 2.

Tabel 1. Hasil Mengelompokkan Data pada Jupyter Python

Provinsi	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	h	width
Aceh	0	0	17	0	0	36	42	54	64	84	108	138	174	216	270	324	378	432	486
Sumatera Utara	402	548	694	840	986	1132	1278	1424	1570	1716	1862	2008	2154	2300	2446	2592	2738	2884	3030
Sulawesi Barat	120	126	132	138	144	150	156	162	168	174	180	186	192	198	204	210	216	222	228
Riau	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400	7400
Jambi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sumatera Selatan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bengkulu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lampung	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300	1300
Kepulauan	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
Bangka Belitung	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DKI Jakarta	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Jawa Barat	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
Jawa Tengah	0	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000	20000
DI Yogyakarta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jawa Timur	300	4500	9000	13500	18000	22500	27000	31500	36000	40500	45000	49500	54000	58500	63000	67500	72000	76500	81000
Banten	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200	2200
Bali	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000	100000
Nusa Tenggara	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Banda	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
Maluku	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Maluku Utara	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sumatera Barat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kalimantan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kalimantan Barat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kalimantan Selatan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kalimantan Timur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kalimantan Tengah	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sulawesi Utara	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sulawesi Tengah	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sulawesi Selatan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sulawesi Tenggara	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Papua	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Papua Barat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3.5. Menampilkan Plot

Implementasi *K-Means* dengan menggunakan *IDE Jupyter* dengan 3 kluster ditampilkan dalam bentuk plot yang dimana pada plot hijau dengan transparan adalah centroid dan yang lainnya adalah data yang dikelompokkan berdasarkan warna yang berbeda-beda. Peluang produksi perikanan dengan implementasi *IDE Jupyter Python* menghasilkan pengelompokan dengan 3 kluster yang terdiri dari 26 provinsi provinsi pada kluster 1 (C1), 3 provinsi pada kluster 2 (C2), dan 5 provinsi pada kluster 3 (C3) Seperti pada Gambar 1.



Gambar 1. Plot Pengelompokan Data Pada Jupyter Python

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian dapat disimpulkan bahwa dengan metode pengelompokan dengan algoritma *K-Means* dapat menyelesaikan masalah pengelompokan data berukuran yang sangat

besar sehingga memudahkan nelayan dan orang-orang yang bekerja dibidang perikanan dengan memberikan informasi tentang provinsi dengan peluang tertinggi untuk budidaya laut.

Hasil pada implementasi *IDE Jupyter Python*, dengan 3 kluster terdiri dari: 26 provinsi pada

kluster 1 (C1) beranggotakan provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan, DKI Jakarta, Jawa Barat, Jawa Tengah, DIYogyakarta, Banten, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, Papua, 3 provinsi pada kluster 2

(C2) beranggotakan provinsi Nusa Tenggara Timur,

Sulawesi Tengah, Sulawesi Selatan, dan 5 provinsi pada kluster 3 (C3) beranggotakan provinsi Jawa Timur, Nusa Tenggara Barat, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Maluku.

Disetiap provinsi menepati kluster yang berbeda-beda. Sehingga disetiap kluster memiliki peluang terhadap provinsi dengan produksi perikanan budidaya laut yang berbeda seperti C1 ialah peluang tertinggi, C2 ialah peluang terendah, dan C3 ialah peluang menengah.

Adapun saran untuk penelitian berikutnya adalah dengan mengelompokkan data dalam skala besar menggunakan metode pengelompokan lainnya. dan perlu dilakukan normalisasi terlebih dahulu, pengolahan dan analisis terhadap data dengan baik sehingga menghasilkan pengelompokan dengan lebih akurat.

Daftar Pustaka

- [1] Sari, R. W., & Hartama, D. (2018, July). Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi. In Seminar Nasional Sains dan Teknologi Informasi (SENSASI) (Vol. 1, No. 1).
- [2] Agusta, Y. (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika, 3(1), 47-60.
- [3] Purnomo, B. S., & Prasetyaningrum, P. T. (2020, Juni). Implementasi Data Mining Dalam Klasifikasi Kunjungan Wisatawan di Kota Yogyakarta Menggunakan Algoritma. JIFTI - Jurnal Ilmiah Teknologi Informasi dan Robotika (Vol. 2, No.1).
- [4] Irfiani, E., & Indriyani, F. (2020). Algoritma K-Means Untuk Pengelompokan Rute Perjalanan

- Wisata Pada Agen Tour & Travel. Indonesian Journal of Computer Science, 9(1), 44-52.
- [5] Maulida, L. (2018). Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means. JISKA (Jurnal Informatika Sunan Kalijaga), 2(3), 167-174.
- [6] Rosiani, U. D., Rahmad, C., Rahmawati, M. A., & Tupamahu, F. (2020). SEGMENTASI BERBASIS K-MEANS PADA DETEKSI CITRA PENYAKIT DAUN TANAMAN JAGUNG. Jurnal Informatika Polinema, 6(3), 37-42.
- [7] Lase, Y., & Panggabean, E. (2019). Implementasi Metode K-Means Pengelompokan Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru. Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP), 2(2), 43-47.
- [8] Hablum, R. J., Khairan, A., & Rosihan, R. (2019). Pengelompokan Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means. JIKO (Jurnal Informatika dan Komputer), 2(1), 26-33.
- [9] Sitanggang, I. S., Risal, A. A. N., & Syaufina, L. (2018, November). Incremental Clustering on Hotspot Data as Forest and Land Fires Indicator in Sumatra. In IOP ConferenceSeries: Earth and Environmental Science (Vol. 187, No. 1, p. 012043). IOP Publishing.
- [10] Sihombing, E. G. (2017). Klasifikasi Data Mining Pada Rumah Tangga Menurut Provinsi Dan Status Kepemilikan Rumah Kontrak/Sewa Menggunakan K-Means Pengelompokan Method. Computer Engineering, Science and System Journal, 2(2), 74-82.