

IMPLEMENTASI K-MEANS CLUSTERING DALAM MENGELOMPOKKAN JUMLAH PENJUALAN IKAN LAUT DI TPI MENURUT WILAYAH

Muhammad Y Rizki¹, Fira Fania², Agus P Windarto³

^{1,2,3}Program Studi Sistem Informasi STIKOM TUNAS BANGSA Pematangsiantar
Email: ryuda1014@gmail.com, firafania23@gmail.com, agus.perdana@amiktunasbangsa.ac.id

(Naskah masuk: 22 Februari 2020, diterima untuk diterbitkan: 23 Maret 2020)

Abstrak

Setiap wilayah di Indonesia memiliki banyak sekali sumber daya alam termasuk produksi ikan, salah satunya adalah ikan laut. Dari sumber daya alam tersebut banyak para nelayan menjadikan hal tersebut menjadi salah satu pekerjaan mereka yaitu menangkap ikan. Ikan hasil tangkapan akan dijual di TPI (Tempat Pelelangan Ikan) yang nantinya akan di timbang dan di harga sesuai dengan standart harga ikan yang berlaku pada saat itu. Penelitian ini membahas tentang "Implementasi K-Means Clustering Dalam mengelompokkan Jumlah Penjualan Ikan DI TPI Menurut Wilayah". Data di dalam penelitian ini peneliti ambil dari sebuah website penyedia data statistic pemerintah yaitu BPS (Badan Pusat Statistik) www.bps.go.id. Terdapat 20 Provinsi di Indonesia di dalam penelitian ini. Peneliti menggunakan algoritma *K-Means clustering* dalam penelitian ini, Peneliti membuat 2 buah *cluster* di dalamnya yakni *cluster* tingkat tinggi (C1) dan *cluster* tingkat rendah (C2). Peneliti memanfaatkan *software RapidMiner* dalam proses pengujian penelitian ini. Hasil dari penelitian ini menghasilkan 2 Provinsi memnduduki posisi *cluster* tingkat tinggi (C1) dan 18 Provinsi menduduki *cluster* tingkat rendah (C2). Penelitian ini diharapkan dapat menjadi informasi kepada pemerintah wilayah Indonesia tentang jumlah penjualan ikan laut di Indonesia serta dapat menjadi acuan dalam menyeimbangkan penjualan ikan di tempat pelelangan atau pun eceran agar dapat menstabilitkan harga ikan.

Kata kunci: data mining, k-means, clustering ikan.

IMPLEMENTATION OF K-MEANS CLUSTERING IN GROUPING THE NUMBER OF SEA FISH SALES IN TPI BY REGION

Abstract

Every region in Indonesia has a lot of natural resources including fish production, one of which is sea fish. From these natural resources a lot of fishermen make it one of their jobs to catch fish. Catch fish will be sold at TPI (Auction Place Fish) which will later be weighed and valued according to the prevailing fish price standards at the time. This study discusses "Implementation of K-Means Clustering in classifying the Number of Fish Sales at TPI by Region". The data in this study the researchers took from a government statistical data provider website, BPS (Central Statistics Agency) www.bps.go.id. There are 20 Provinces in Indonesia in this study. Researchers used the K-Means clustering algorithm in this study, the researchers made 2 clushers in it namely high level cluster (C1) and low level cluster (C2). Researchers utilize RapidMiner software in the process The results of this study resulted in 2 provinces having a high level cluhter position (C1) and 18 provinces occupying a low level cluster (C2). This research is expected to provide information to the Indonesian regional government about the amount of sea fish sales in Indonesia and can be a reference in balancing fish sales at the auction or retail in order to stabilize fish pices.

Keywords: data mining, k-means, clustering fish.

1. PENDAHULUAN

Setiap wilayah di Indonesia memiliki banyak sekali sumber daya alam termasuk produksi ikan, salah satunya adalah ikan laut. Dari sumber daya alam tersebut banyak para nelayan menjadikan hal tersebut menjadi salah satu pekerjaan mereka yaitu menangkap ikan. Ikan hasil tangkapan akan dijual di

TPI (Tempat Pelelangan Ikan) yang nantinya akan di timbang dan di harga sesuai dengan standart harga ikan yang berlaku pada saat itu. Penelitian ini membahas tentang implementasi k-means clustering dalam mengelompokkan jumlah penjualan ikan di tpi menurut wilayah. Implementasi k-means pernah dilakukan oleh [1] pada data siswa. [2] melakukan pembentukan cluster pada database. K-means juga

dinilai dapat menyelesaikan berbagai permasalahan terkait [3]. sebagaimana [4] menentukan strategi marketing.

Pada penelitian sebelumnya [8] membahas penerapan algoritma *clustering* dalam mengelompokkan banyaknya desa/kelurahan menurut upaya antisipasi/ mitigasi bencana alam menurut provinsi dengan k-means. Penelitian tersebut mampu menghasilkan model pengelompokan dari wilayah-wilayah di Indonesia berdasarkan upaya antisipasi bencana alam sehingga mampu menghasilkan informasi dan masukan kepada pemerintah dalam kasus tersebut. Sedangkan [11] melakukan klaster hasil tangkap ikan di Ternate.

Penelitian sebelumnya [6] membahas tentang Dampak Pelelangan Terhadap Stabilisasi Harga Ikan pada Tingkat Produsen di Pantai Utara Jawa. Berdasarkan paper tersebut pada penjualan sistem lelang harga ikan lebih stabil dari harga yang tidak menggunakan sistem lelang dengan kata lain sistem lelang dapat menstabilkan harga pada tingkat produsen.[6]

Data di dalam penelitian ini peneliti ambil dari sebuah website penyedia data statistik pemerintah yaitu BPS (Badan Pusat Statistik) www.bps.go.id. Terdapat 20 Provinsi di Indonesia di dalam penelitian ini. Peneliti menggunakan algoritma K-Means *clustering* dalam penelitian ini, Peneliti membuat 2 buah cluster di dalamnya yakni cluster tingkat tinggi(C1) dan cluster tingkat rendah(C2). Peneliti memanfaatkan *software RapidMiner* dalam proses pengujian penelitian ini. Hasil dari penelitian ini menghasilkan 2 Provinsi menduduki posisi cluster tingkat tinggi (C1) dan 18 Provinsi menduduki cluster tingkat rendah (C2). Penelitian ini diharapkan dapat menjadi informasi kepada pemerintah wilayah Indonesia tentang jumlah penjualan ikan laut di Indonesia serta dapat menjadi acuan dalam menyeimbangkan penjualan ikan di tempat pelelangan ataupun eceran agar dapat menstabilkan harga ikan seperti yang dijelaskan pada peneliti sebelumnya.

2. METODE PENELITIAN

Data mining merupakan salah metode yang saat ini banyak diterapkan diberbagai macam bidang keilmuan. Hal ini dinilai sangat efektif dalam menganalisis suatu data tertentu. Tak terkecuali dalam data-data perikanan di Indonesia. Dalam penelitian ini metode yang digunakan yaitu metode Data Mining sebagai berikut [7]

- (a) Tahap pengumpulan data,
- (b) Tahap *clustering*, dan
- (c) Tahap analisis

A. Tahap Pengumpulan Data

Dalam penelitian ini sumber data di dalam penelitian diperoleh dari sebuah website yang menyediakan data statistik yakni www.bps.go.id. Data terdiri dari 20 Provinsi di Indonesia. Terdapat 2 cluster yakni tingkat tinggi (C1) dan cluster tingkat rendah(C2).

C. Tahap *Clustering*

Pada tahapan ini akan dilakukan proses perhitungan jarak antar *cluster* dengan menentukan bobot-bobot nilai centroid data di dalam data tersebut yang nantinya akan dilakukan proses analisis lebih lanjut untuk mendapatkan hasil akhir yang sesuai. [8] – [10]. Dalam penelitian ini juga untuk penentuan kedekatan data menggunakan *euclidean distance*.

D. Tahap Analisis

Selanjutnya data akan dianalisis menggunakan *K-Means Clustering* dengan menghitung nilai-nilai cluster dari setiap atribut dan menentukan jarak terpendek dari setiap atribut dan selanjutnya adalah membentuk pola cluster.

3. HASIL DAN PEMBAHASAN

Sebagaimana telah dijelaskan pada metode penelitian bahwa tahapan awal dari penelitian ini adalah tahap pengumpulan data. Berikut adalah data dari bps dari tahun 2015-2018 dapat dilihat pada Tabel 1 berikut:

Tabel 1. Data BPS Produksi Perikanan

NO	Provinsi	produksi perikanan laut yang dijual di TPI (Tempat Pelelangan Ikan) 2015-2018 (Ton)			
		2015	2016	2017	2018
1	ACEH	45190,9	21479,5	43305,88	43709,09
2	SUMATERA UTARA	3965,3	4571,7	15436,76	20570,57
3	SUMATERA BARAT	1709,7	666,9	281,9	1792,09
4	BENGKULU	2615,4	2565,4	3441,58	2012,91
5	LAMPUNG	4105,9	1159,2	1533,14	3276,78
6	KEP. BANGKA BELITUNG	1242,4	2936,6	4755,81	6413,07
7	DKI JAKARTA	24035,1	104956,9	31228,34	103835,32
8	JAWA BARAT	52361,5	46470,2	44613,42	44547,91

9	DI YOGYAKARTA	2339,5	1700	3088,7	1829,37
10	JAWA TIMUR	71696,8	59520,2	156599,68	114811,2
11	BANTEN	5401,3	5758,4	6474,66	6496,03
12	BALI	11153,2	8664,5	8451,25	18256,3
13	NUSA TENGGARA BARAT	6122,6	3530,5	3185,07	5181,95
14	KALIMANTAN BARAT	4159,2	3838,1	2932,12	8568,37
15	KALIMANTAN TIMUR	12433,4	18694,2	14336,18	19544,23
16	SULAWESI UTARA	9595	15306,1	51908,9	61094,61
17	SULAWESI TENGAH	3455,3	4149,5	3409,3	2373,26
18	SULAWESI SELATAN	29684,7	20601,5	45546,4	46713,97
19	SULAWESI TENGGARA	4161,8	5920,4	6678,31	18572,22
20	GORONTALO	4383,4	3188,8	3219,02	3530,67

Selanjutnya adalah menentukan *centroid data* awal dari data yang telah diperoleh ke dalam 2 *cluster* yakni *cluster* tingkat tinggi (C1) dan *cluster* tingkat rendah (C2).

Tabel 2. Centroid Data Awal

	a	b	c	d
c1	71696,8	104956,9	156599,7	114811,2

c2	1242,4	666,9	281,9	1792,09
----	--------	-------	-------	---------

Keterangan :
A = 2015
B = 2016
C = 2017
D = 2018

Tabel 3. Proses Iterasi 1

NO	Provinsi	Produksi perikanan laut yang dijual di TPI (Tempat Pelelangan Ikan) 2015-2018 (Ton)				c1	c2	jp
		2015	2016	2017	2018			
1	ACEH	45190,9	21479,5	43305,88	43709,09	159881,3	77283,46	77283,46
2	SUMATERA UTARA	3965,3	4571,7	15436,76	20570,57	208501,8	24595,99	24595,99
3	SUMATERA BARAT	1709,7	666,9	281,9	1792,09	230180,7	467,3	467,3
4	BENGKULU	2615,4	2565,4	3441,58	2012,91	226797,5	3939,768	3939,768
5	LAMPUNG	4105,9	1159,2	1533,14	3276,78	227657,7	3494,553	3494,553
6	KEP. BANGKA BELITUNG	1242,4	2936,6	4755,81	6413,07	224006	6820,621	6820,621
7	DKI JAKARTA	24035,1	104956,9	31228,34	103835,32	134573,7	150885,4	134573,7
8	JAWA BARAT	52361,5	46470,2	44613,42	44547,91	145850,6	92219,57	92219,57
9	DI YOGYAKARTA	2339,5	1700	3088,7	1829,37	227602,6	3185,976	3185,976
10	JAWA TIMUR	71696,8	59520,2	156599,68	114811,2	45436,7	213626,1	45436,7
11	BANTEN	5401,3	5758,4	6474,66	6496,03	220238,8	10183,18	10183,18
12	BALI	11153,2	8664,5	8451,25	18256,3	210258,3	22360,55	22360,55
13	NUSA TENGGARA BARAT	6122,6	3530,5	3185,07	5181,95	223923,9	7206,671	7206,671
14	KALIMANTAN BARAT	4159,2	3838,1	2932,12	8568,37	222907,1	8456,114	8456,114
15	KALIMANTAN TIMUR	12433,4	18694,2	14336,18	19544,23	200669,1	31030,36	31030,36
16	SULAWESI UTARA	9595	15306,1	51908,9	61094,61	160435,4	80412,74	80412,74
17	SULAWESI TENGAH	3455,3	4149,5	3409,3	2373,26	225673,7	5209,973	5209,973
18	SULAWESI SELATAN	29684,7	20601,5	45546,4	46713,97	160782,2	72616,82	72616,82
19	SULAWESI TENGGARA	4161,8	5920,4	6678,31	18572,22	214726,7	18936,97	18936,97
20	GORONTALO	4383,4	3188,8	3219,02	3530,67	225381,5	5279,696	5279,696

Tabel 4. Hasil Proses Iterasi 1

NO	Provinsi	c1	c2
1	ACEH		1
2	SUMATERA UTARA		1
3	SUMATERA BARAT		1
4	BENGKULU		1
5	LAMPUNG		1
	KEP. BANGKA		
6	BELITUNG		1
7	DKI JAKARTA	1	
8	JAWA BARAT		1
9	DI YOGYAKARTA		1
10	JAWA TIMUR	1	
11	BANTEN		1
12	BALI		1
	NUSA TENGGARA		
13	BARAT		1
14	KALIMANTAN BARAT		1
15	KALIMANTAN TIMUR		1
16	SULAWESI UTARA		1
17	SULAWESI TENGAH		1
18	SULAWESI SELATAN		1
	SULAWESI		
19	TENGGARA		1
20	GORONTALO		1

Proses iterasi ke-1 terdapat dua data di *cluster* 1 atau kategori tinggi yaitu DKI Jakarta dan Jawa Timur sedangkan 18 Data lainnya berada pada *cluster* 2 atau kategori rendah. Proses dilanjutkan pada iterasi kedua. Dengan nilai centroid yang digunakan adalah pada Tabel 5. Proses iterasi akan terus berlanjut sampai hasil dari proses iterasi ke-n sama dengan hasil proses iterasi sebelumnya atau ketika posisi data dalam cluster sudah tidak mengalami perubahan atau perindahan dari satu klaster ke klaster yang lain.

Tabel 5. Centroid Data Iterasi 2

	a	b	c	d
c1	47865,95	82238,55	93914,01	109323,3
c2	11337,81	9511,194	14588,8	17471,3

Setelah dilakukan perhitungan pada iterasi kedua, dengan mengukur kedekatan data menggunakan euclidean distance proses dari iterasi kedua ditunjukkan pada Tabel 6. dengan hasil dari iterasi kedua ditunjukkan pada Tabel 7.

Tabel 7. Hasil Proses Iterasi 2

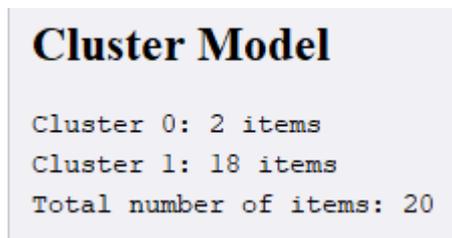
NO	Provinsi	c1	c2
1	ACEH		1
2	SUMATERA UTARA		1
3	SUMATERA BARAT		1
4	BENGKULU		1
5	LAMPUNG		1
	KEP. BANGKA		
6	BELITUNG		1
7	DKI JAKARTA	1	
8	JAWA BARAT		1
9	DI YOGYAKARTA		1
10	JAWA TIMUR	1	
11	BANTEN		1
12	BALI		1
	NUSA TENGGARA		
13	BARAT		1
	KALIMANTAN		
14	BARAT		1
	KALIMANTAN		
15	TIMUR		1
16	SULAWESI UTARA		1
17	SULAWESI TENGAH		1
18	SULAWESI SELATAN		1
	SULAWESI		
19	TENGGARA		1
20	GORONTALO		1

Pada literasi kedua sudah terlihat bahwa sudah tidak ada data yang mengalami perubahan posisi *cluster* atau klaster sebelumnya telah sama dengan klaster ke-n sehingga proses iterasi berhenti. Dengan demikian hanya terjadi sebanyak 2 kali iterasi pada data-data produksi perikanan laut yang dijual di Tempat Pelelangan Ikan (TPI). Hal ini disebabkan jumlah data dalam pengujian yang terbatas sehingga iterasi yang terjadi saat penerapan algoritma k-means *clustering* juga tidak banyak terjadi. Provinsi yang berada didalam cluster satu atau kategori tingkat tinggi yaitu provinsi DKI Jakarta dan Jawa Timur. Sedangkan 18 Provinsi lainnya berada di cluster 2 kategori tingkat rendah yaitu Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, Lampung, Kepulauan Bangka Belitung, Jawa Barat, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatanm Sulawesi Tenggara dan Gorontalo.

Tabel 6. Proses Iterasi 2

NO	Provinsi	Produksi perikanan laut yang dijual di TPI (Tempat Pelelangan Ikan) 2015-2018(Ton)				c1	c2	jp
		2015	2016	2017	2018			
1	ACEH	45190,9	21479,5	43305,88	43709,09	102787,3	52937,37	52937,37
2	SUMATERA UTARA	3965,3	4571,7	15436,76	20570,57	148307,5	9438,059	9438,059
3	SUMATERA BARAT	1709,7	666,9	281,9	1792,09	170629	24928,84	24928,84
4	BENGKULU	2615,4	2565,4	3441,58	2012,91	167619,2	22080,46	22080,46
5	LAMPUNG	4105,9	1159,2	1533,14	3276,78	168133,5	22225,91	22225,91
	KEP. BANGKA							
6	BELITUNG	1242,4	2936,6	4755,81	6413,07	164323,6	19081,79	19081,79
7	DKI JAKARTA	24035,1	104956,9	31228,34	103835,32	71018,61	130409,8	71018,61
8	JAWA BARAT	52361,5	46470,2	44613,42	44547,91	89027,97	68436,34	68436,34
9	DI YOGYAKARTA	2339,5	1700	3088,7	1829,37	168414	22779,52	22779,52
10	JAWA TIMUR	71696,8	59520,2	156599,68	114811,2	71018,61	189172,6	71018,61
11	BANTEN	5401,3	5758,4	6474,66	6496,03	160846,3	15349,96	15349,96
12	BALI	11153,2	8664,5	8451,25	18256,3	149526,1	6247,937	6247,937
	NUSA TENGGARA							
13	BARAT	6122,6	3530,5	3185,07	5181,95	164361,2	18548,32	18548,32
14	KALIMANTAN BARAT	4159,2	3838,1	2932,12	8568,37	162745,7	17287,47	17287,47
15	KALIMANTAN TIMUR	12433,4	18694,2	14336,18	19544,23	140307,7	9480,969	9480,969
16	SULAWESI UTARA	9595	15306,1	51908,9	61094,61	100175,1	57726,96	57726,96
17	SULAWESI TENGAH	3455,3	4149,5	3409,3	2373,26	166432	21066,89	21066,89
18	SULAWESI SELATAN	29684,7	20601,5	45546,4	46713,97	101926,6	47677,15	47677,15
19	SULAWESI TENGGARA	4161,8	5920,4	6678,31	18572,22	153558,9	11321,52	11321,52
20	GORONTALO	4383,4	3188,8	3219,02	3530,67	166003,8	20296,54	20296,54

Selanjutnya dapat menggunakan bantuan software RapidMiner dalam proses pengujian penelitian ini. Hasil proses pengujian dapat dilihat pada gambar 1,2,3,dan 4 berikut:



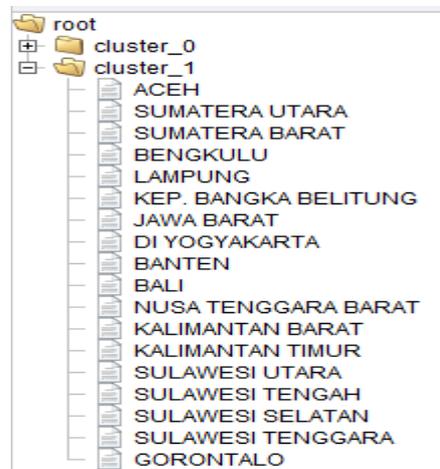
Gambar 1. Cluster Model

Attribute	cluster_0	cluster_1
2015.0	47865.950	11337.806
2016.0	82238.550	9511.194
2017.0	93914.010	14588.800
2018.0	109323.260	17471.300

Gambar 2. Centroid Data Iterasi 2



Gambar 3. Anggota Cluster Tinggi (C1)



Gambar 4. Anggota Cluster Rendah (C2)

4. KESIMPULAN

Algoritma K-Means dapat menjadi solusi dalam kasus penelitian ini yakni menentukan jumlah penjualan ikan laut menurut wilayah. Terdapat 2 Provinsi yang menduduki posisi *cluster* tinggi yakni DKI Jakarta dan Jawa Timur. Sedangkan 18 Provinsi lainnya menduduki *cluster* rendah. Penelitian ini diharapkan dapat menjadi acuan pemerintah wilayah dalam menstabilkan harga ikan terutama untuk provinsi yang menduduki *cluster* tingkat tinggi. *Software RapidMiner* dapat menjadi salah satu solusi dalam proses pengujian algoritma *K-Means clustering*. Selain itu perlu juga diterapkan validasi algoritma clustering, sehingga bisa didapat cluster terbaik pada data-data perikanan di Indonesia.

5. DAFTAR PUSTAKA

- [1] N. Butarbutar., A. Perdana., D. Windarto dan M. Hartama., 2016., “Komparasi Kinerja Algoritma Fuzzy C-Means Dan K-Means Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai akademik Siswa (Studi Kasus : SMP Negeri 2 Pematangsiantar),” *J. Ris. Sist. Inf. Tek. Inform.*, vol. 1, no. 1, 46-55.
- [2] S. Andayani., 2016. “K-Pembentukan cluster dalam Knowledge Discovery in Database dengan Algoritma K- means,” no. June.
- [3] Y. Agusta., 2017., “K-Means – Penerapan, Permasalahan dan Metode Terkait,” *J. Sist. dan Inform.*, vol. 3, no. Februari, pp. 47–60.
- [4] J. O. Ong., 2013., “Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university,” *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20.
- [5] V. T. Manurung and M. Syukur., 2016., “Dampak Pelelangan Terhadap Stabilisasi Harga Ikan pada Tingkat Produsen di Pantai Utara Jawa,” *Forum Penelit. Agro Ekon.*, vol. 7, no. 2, p. 12, 2016, doi: 10.21082/fae.v7n2.1989.12-19.
- [6] M. G. Sadewo., A. P. Windarto and A. Wanto., 2018. “Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, doi: 10.30865/komik.v2i1.943.
- [7] F. Nasari and S. Darma., 2015., Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Potensi Utama), “*Seminar Nasional Teknologi Informasi dan Multimedia*” pp. 6–8.
- [8] E. Muningsih and S. Kiswati., 2015., “Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang,” *J. Bianglala Inform.*, vol. 3, no. 1, pp. 10–17.
- [9] Y. Darmi and A. Setiawan., 2016., “Penerapan Metode Clustering K-Means Dalam,” *Y. Darmi, A. Setiawan*, vol. 12, no. 2, pp. 148–157.
- [10] T. Khotimah., 2014., “Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 5, no. 1, pp. 83–88, 2014, doi: 10.24176/simet.v5i1.141.
- [11] R. J. Hablum., A. Khairan dan R. Rosihan., 2019., “Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (PPN) Ternate Menggunakan Algoritma K-Means,” *JIKO (Jurnal Informatika dan Komputer)*., vol. 2, no. 1, pp. 26-33.