

Penentuan *Cluster* Optimum pada Tingkat Pengangguran dan Tingkat Kemiskinan di Jawa Timur Menggunakan Algoritma *Fuzzy C-Means*

Ari Eko Wardoyo¹, Nigati Triuspita²

¹Jurusan Manajemen Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember, arieko@unmuhjember.ac.id

²Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember, nitritaoke@gmail.com

Keywords:

Clustering Analysis, Optimization, Fuzzy C-Means, Unemployment Data, Poverty Data,

ABSTRACT (10 PT)

There are many methods used in resolving data clustering. One of them is the Fuzzy C-Means (FCM) method, which is a reliable method to solve clustering problems in the East Java region. This study aims to determine the optimum cluster in the East Java region which can help the government to identify problems and assist policymaking in regencies/cities in East Java province. The research process uses data from the central statistical agency, namely the unemployment rate and poverty rate from 2010 to 2015. In this study, the Davies Bouldin Index (DBI) is used as a cluster validation test for determining the optimum cluster. Unemployment rate and poverty rate data were analyzed using RStudio. From the calculation of the FCM method and also the determination of the optimum cluster results obtained in 2 clusters with a DBI value of 1.2759, 3 clusters with a DBI value of 0.9937, 4 clusters with a DBI value of 0.8737. The optimum cluster is in 4 clusters with a minimum DBI value.

Kata Kunci

Analisis Pengelompokan, Optimasi, Fuzzy C-Means, Data Pengangguran, Data Kemiskinan,

ABSTRAK (10 PT)

Ada banyak metode yang digunakan dalam menyelesaikan *clustering* suatu data. Salah satunya adalah metode *Fuzzy C-Means* (FCM) yang merupakan salah satu metode yang dapat diandalkan dalam menyelesaikan masalah *clustering* di wilayah Jawa Timur. Penelitian ini bertujuan untuk menentukan *cluster* optimum di wilayah Jawa Timur yang dapat membantu pemerintah untuk mengetahui permasalahan dan membantu pengambilan kebijakan pada wilayah kabupaten/kota di Provinsi Jawa Timur. Proses penelitian menggunakan data dari badan pusat statistik yaitu tingkat pengangguran dan tingkat kemiskinan pada tahun 2010 sampai 2015. Pada penelitian ini digunakan *Davies Bouldin Index* (DBI) sebagai uji validasi *cluster* untuk penentuan *cluster* optimum. Data tingkat pengangguran dan tingkat kemiskinan di analisis menggunakan *RStudio*. Dari hasil perhitungan metode FCM dan juga penentuan *cluster* optimum didapat hasil pada 2 *cluster* dengan nilai DBI 1.2759, 3 *cluster* dengan nilai DBI 0.9937, 4 *cluster* dengan nilai DBI 0.8737. *Cluster* optimum terdapat pada 4 *cluster* yang ditandai dengan nilai DBI yang paling kecil.

Korespondensi Penulis:

Ari Eko Wardoyo,
Universitas Muhammadiyah Jember,
Jalan Karimata 49 Jember
Email: arieko@unmuhjember.ac.id

1. PENDAHULUAN

Upaya menurunkan tingkat pengangguran dan menurunkan tingkat kemiskinan adalah sama pentingnya. Pengangguran dan kemiskinan adalah sebuah contoh permasalahan pembangunan dan sosial kemasyarakatan. Pembangunan sendiri merupakan aspek penting untuk melihat kinerja seberapa efektifnya penggunaan sumber daya yang ada sehingga lapangan kerja dapat menyerap angkatan kerja yang tersedia. Semakin meningkat produksi barang dan jasa maka semakin banyak juga angkatan kerja yang dibutuhkan sehingga tingkat pengangguran dan kemiskinan semakin menurun. Secara teori, masyarakat yang mempunyai pekerjaan dan berpenghasilan serta dapat memenuhi kebutuhan hidupnya, maka dapat dikatakan masyarakat yang tidak miskin.

Jumlah pengangguran akan meningkat dan menurun setiap tahunnya sesuai dengan pertumbuhan penduduk. Menurut data Badan Pusat Statistik (BPS), di Jawa timur sendiri tingkat pengangguran pada tahun 2015

sekitar 4,47%, pada tahun 2016 menurun sekitar 4,21% dan pada tahun 2012 menurun sekitar 4% [1]. Untuk tingkat kemiskinan pada tahun 2015 sekitar 12,34%, pada tahun 2016 menurun sekitar 12,05% dan pada tahun 2017 menurun sekitar 11,77% [2]. Jika dilihat per kabupatennya pada tahun 2017 pengangguran tertinggi yaitu di Kota Malang sekitar 7,22% sedangkan kemiskinan tertinggi yaitu kabupaten Sampang sekitar 23,56% dibanding kabupaten/kota lainnya [3].

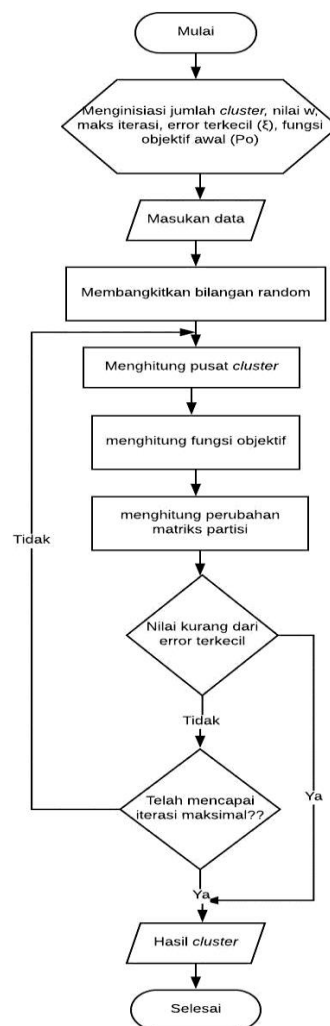
Pada penelitian sebelumnya [4]–[6] sudah ada pengetahuan untuk mengetahui kelompok wilayah namun menggunakan atribut atau variabel yang berbeda. Penelitian ini mengelompokkan wilayah berdasarkan tingkat pengangguran terbuka dan tingkat kemiskinan yang dapat membantu pemerintah untuk mengetahui permasalahan dan membantu pengambilan kebijakan wilayah kabupaten/kota di Provinsi Jawa Timur. Penentuan wilayah kota yang memiliki tingkat pengangguran yang tinggi sangat berpengaruh pada tingkat kemiskinan berdasarkan pengetahuan dari jumlah data Badan Pusat Statistik. Karena hal itu, untuk mendapatkan pengetahuan tersebut maka diperlukan suatu proses pengolahan data dengan teknik Data Mining. Teknik Data Mining yang akan digunakan pada penelitian ini menggunakan metode *Fuzzy C-Means* [7]. Penelitian diharapkan dapat menghasilkan suatu informasi yang dapat meng-*cluster* atau mengelompokkan sebuah wilayah kabupaten/kota di Jawa Timur.

2. METODE PENELITIAN

Metode penelitian dilakukan dengan pendekatan kuantitatif dimana analisisnya berupa data-data *numeric* (angka). Penelitian ini menggunakan metode Algoritma *Fuzzy C-Means* yang berguna untuk memartisi data berdasarkan pada jarak antara data masukkan dengan pusat kluster terdekat.

2.1 Penerapan Fuzzy C-Means

Diagram alur dari metode *Fuzzy C-Means* yang digunakan berdasarkan pengaruh tingkat pengangguran terhadap tingkat kemiskinan di Jawa Timur, pada umumnya kinerja *Fuzzy C-Means* adalah sebagai berikut :



Gambar 1. Flowchart Fuzzy C-Means

Langkah-langkah dalam proses pengelompokan data menggunakan algoritma *Fuzzy C-Means* adalah:

- Menetapkan Matriks partisi awal (U), berupa data tingkat pengangguran dan data tingkat kemiskinan selama enam tahun terakhir. Parameter yang digunakan adalah mulai tahun 2010-2015 dengan menggunakan 10 data sampel. Detail data yang digunakan sebagai berikut:

Data ke	Kabupaten/kota	TPT 2010	TK 2010	TPT 2011	TK 2011	TPT 2012	TK 2012	TPT 2013	TK 2013	TPT 2014	TK 2014	TPT 2015	TK 2015
1	Trenggalek	2.15	15.98	3.27	14.9	2.98	14.21	4.04	13.56	4.2	13.1	2.46	13.39
2	Kediri	3.75	15.52	8.33	14.4	4.08	13.71	4.65	13.23	4.91	12.77	5.02	12.91
3	Lumajang	3.17	13.98	3.16	13	4.6	12.4	2.01	12.14	2.83	11.75	2.6	11.52
4	Banyuwangi	3.92	11.25	6.06	10.5	3.41	9.97	4.65	9.61	7.17	9.29	2.55	9.17
5	Probolinggo	2.02	25.22	2.8	23.5	1.92	22.22	3.3	21.21	1.47	20.44	2.51	20.82
6	Mojokerto	4.84	12.23	6.79	11.4	3.35	10.71	3.16	10.99	3.81	10.56	4.05	10.57
7	Bojonegoro	3.29	18.78	5.7	17.5	3.42	16.66	5.81	16.02	3.21	15.48	5.01	15.71
8	Sampang	1.77	32.47	2.13	30.2	1.71	27.97	4.68	27.08	2.22	25.8	2.51	25.69
9	Kota Malang	8.68	5.9	9.74	5.5	7.96	5.21	7.73	4.87	7.22	4.8	7.28	4.6
10	Kota Surabaya	6.84	7.07	7.81	6.6	5.27	6.25	5.32	6	5.82	5.79	7.01	5.82

Gambar 2. Data Tingkat Pengangguran dan Tingkat Kemiskinan

Keterangan:

- TPT: Tingkat Pengangguran Terbuka
- TK: Tingkat Kemiskinan

- Menentukan nilai parameter awal

Tabel 1. Parameter Awal

No	Keterangan		Nilai
1	Jumlah cluster	c	3
2	Pangkat	w	2
3	Maksimum Iterasi	MaxIter	20
4	Error terkecil	ξ	0,001
5	Fungsi Objektif Awal	P ₀	0
6	Iterasi Awal	t	1
7	Fungsi Objektif	P1	5.545,94
8	P _t - P _{t-1}		5.545,94

- Membangkitkan bilangan *random* sebagai data derajat keanggotaan awal untuk elemen matriks partisi awal (U). Berikut hasilnya:

Tabel 2. Bilangan Random

No	C1	C2	C3
1	0,441	0,490	0,069
2	0,556	0,385	0,059
3	0,156	0,307	0,537
4	0,461	0,075	0,464
5	0,427	0,216	0,357
6	0,096	0,285	0,619
7	0,329	0,399	0,272
8	0,020	0,467	0,514
9	0,311	0,348	0,341
10	0,321	0,538	0,140

- Menentukan pusat *cluster* pada iterasi awal

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

$$V_{kj} = \frac{((0,441)^2 * 2,15)}{0,441^2}$$

$$V_{kj} = \frac{0,194481 * 2,15}{0,194481} = 2,15$$

5. Menghitung fungsi objektif (P)

$$P_{1.1} = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w$$

$$P_{1.1} = ((2,15 - 3,8734)^2 + (15,98 - 15,0629)^2 + (3,27 - 6,0587)^2 + (14,9 - 14,0256)^2 + (2,98 - 3,8247)^2 + (14,21 - 13,3242)^2 + (4,04 - 3,7024)^2 + (13,56 - 12,7862)^2 + (4,2 - 4,7390)^2 + (13,1 - 12,3504)^2 + (2,46 - 4,1135)^2 + (13,39 - 12,4684)^2) * (0,441)^2$$

$$P_{1.1} = 3,6950$$

$$P_{tot} = P_{1.1} + P_{1.2} + P_{1.3}$$

$$P_{tot} = 3,6950 + 3,5733 + 0,0612 = 7,3294$$

Tabel 3. Fungsi Objektif (P) Iterasi 1

Data	P-1.1	P-1.2	P-1.3	P Tot
1	3,6950	3,5733	0,0612	7,3295
2	3,3803	4,0541	4,6292	12,0636
3	1,0685	93,7222	287,4253	382,2160
4	157,7353	4,2084	159,5884	321,5321
5	549,7745	141,3352	384,2248	1075,3345
6	7,9084	69,8954	329,4847	407,2886
7	195,3573	286,5817	133,8407	615,7797
8	1,8637	1057,5684	1279,2094	2338,6415
9	54,1647	67,4899	64,9133	186,5679
10	50,0041	139,6609	9,5185	199,1835
Fungsi Objektif (P1)				5545,9367

6. Menghitung perubahan matriks partisi

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}$$

$$\mu_{ik} = \frac{3,6950}{7,3295} = 0,5041$$

Tabel 4. Derajat Keanggotaan Iterasi 1

Data	μ_1	μ_2	μ_3	C1	C2	C3
1	0,5041	0,4875	0,0084	OK		
2	0,2802	0,3361	0,3837			OK
3	0,0028	0,2452	0,7520			OK
4	0,4906	0,0131	0,4963			OK
5	0,5113	0,1314	0,3573	OK		
6	0,0194	0,1716	0,8090			OK
7	0,3173	0,4654	0,2174		OK	
8	0,0008	0,4522	0,5470			OK
9	0,2903	0,3617	0,3479		OK	
10	0,2510	0,7012	0,0478		OK	
Total				2	3	5

7. Memeriksa kondisi berhenti

Iterasi akan berhenti apabila memenuhi syarat berikut:

- Jika ($|P_t - P_{t-1}| < \xi$) atau ($t > MaxIter$) maka berhenti, artinya jika hasil pengurangan dari fungsi objektif dan fungsi objektif awal kurang dari *error* terkecil atau iterasi awal lebih dari maksimal iterasi maka akan berhenti
- Jika tidak $t = t+1$ maka mengulang kembali tahap/langkah ke empat

Tabel 5. Ketentuan Berhenti

Ketentuan Berhenti	
Jika ($ P_t - P_{t-1} < \xi$)	Lanjut
($t > MaxIter$)	Lanjut

Karena $|P_t - P_{t-1}| = |5,545,94 - 0| = 5,545,94 > error$ terkecil maka tidak berhenti dan dilanjutkan pada iterasi selanjutnya hingga $P_t - P_{t-1} < error$ terkecil atau iterasi awal lebih dari maksimal iterasi. Pada proses uji coba ini berhenti pada iterasi 14.

2.2 Proses Davies Bouldin Index

Penentuan *cluster* optimum pada penelitian ini menggunakan *Davies-Bouldin Index*, dengan jumlah 10 data sampel dengan hasil terbaik.

1. Mengelompokkan data berdasarkan *cluster* yang diikuti Menghitung DBI pada *Fuzzy C-Means* menggunakan hasil dari pusat *cluster* yang kemudian dikelompokkan sesuai dengan hasil derajat keanggotaan
2. Menentukan nilai perhitungan kerapatan (*density*) data pada *cluster* atau *Sum of Square Within (SSW)*
3. Melakukan perhitungan untuk nilai separasi atau *Sum of Square Between (SSB)*
4. Menghitung nilai rasio yang diperoleh

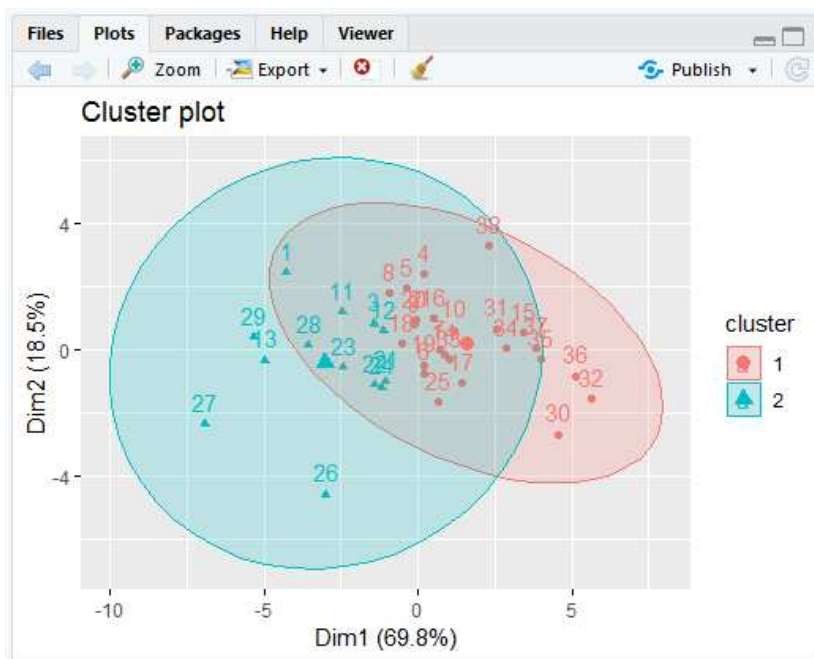
3. HASIL DAN ANALISIS

Penerapan *clustering* dengan *Fuzzy C-Means* akan dibahas secara runtut pada bab ini. Data yang telah diperoleh akan diolah untuk mendapatkan hasil *cluster* optimum menggunakan *Davies Bouldin Index*. Data yang digunakan adalah data tingkat pengangguran terbuka dan tingkat kemiskinan di Provinsi Jawa Timur Periode 2010-2015.

3.1 Penerapan Fuzzy C-Means

- a. Dua Cluster

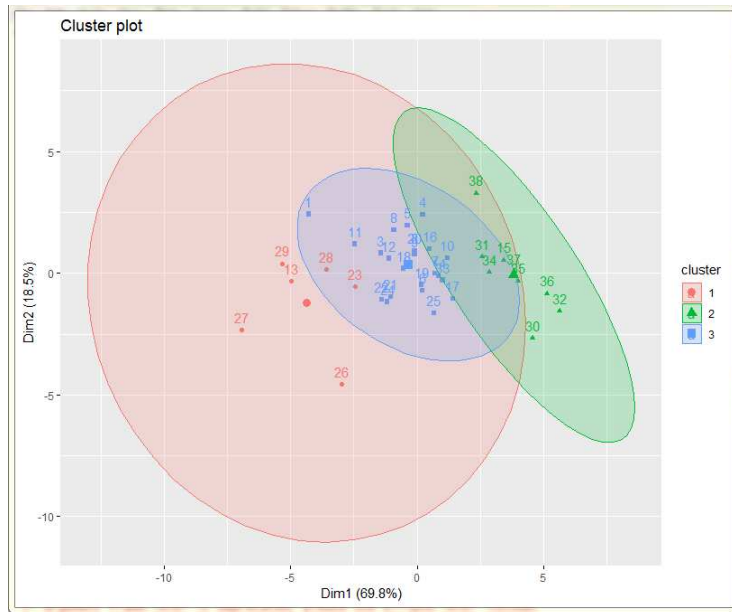
Perhitungan untuk 2 *cluster* berhenti pada iterasi ke-70 dan *error* terkecil adalah 0,001.



Gambar 3. Grafik Plot Untuk 2 Cluster

- b. Tiga Cluster

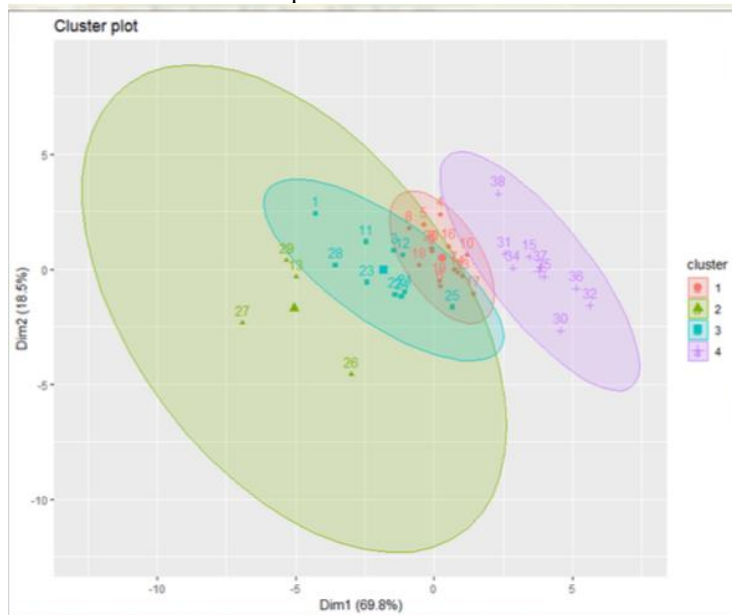
Perhitungan untuk 3 *cluster* berhenti pada iterasi ke-64 dan *error* terkecil adalah 0,001.



Gambar 4. Grafik Plot Untuk 3 Cluster

c. Empat Cluster

Perhitungan untuk 4 cluster berhenti pada iterasi ke-68 dan error terkecil adalah 0,001.



Gambar 5. Grafik Plot Untuk 4 Cluster

3.2 Penentuan Cluster Optimum

Berdasarkan ketiga jumlah cluster (2 cluster, 3 cluster, dan 4 cluster) maka akan dicari nilai cluster optimum.

1. Proses *Davies Bouldin Index* untuk 2 Cluster

- Menentukan nilai perhitungan kerapatan (*density*) data pada cluster atau *Sum of Square Within* (SSW)

Tabel 6. SSW Untuk 2 Cluster

SSW ₁	1867,114
SSW ₂	1441,282

- Mencari nilai separasi atau *Sum of Square Between* (SSB)
SSB = 4257,61
- Nilai rasio yang diperoleh
DBI = 1,2759

2. Proses *Davies Bouldin Index* untuk 3 Cluster
 - Menentukan nilai perhitungan kerapatan (*density*) data pada cluster atau *Sum of Square Within* (SSW)

Tabel 7. SSW Untuk 3 Cluster

SSW ₁	480,1896
SSW ₂	223,8220
SSW ₃	102,2332

- Mencari nilai separasi atau *Sum of Square Between* (SSB)
SSB = 5475,362
 - Nilai rasio yang diperoleh
DBI = 0,9937
3. Proses *Davies Bouldin Index* untuk 4 Cluster
 - Menentukan nilai perhitungan kerapatan (*density*) data pada cluster atau *Sum of Square Within* (SSW)

Tabel 8. SSW Untuk 3 Cluster

SSW ₁	353,3696
SSW ₂	217,1700
SSW ₃	290,5308
SSW ₄	223,8220

- Mencari nilai separasi atau *Sum of Square Between* (SSB)
SSB = 6217,408
 - Nilai rasio yang diperoleh
DBI = 0,8737
4. Hasil DBI Untuk Mencari Cluster Optimum
Semakin kecil nilai *Davies Bouldin Index* yang diperoleh maka semakin baik cluster yang diperoleh dari pengelompokan. Maka untuk cluster optimum pada tingkat pengangguran dan tingkat kemiskinan di Jawa Timur menggunakan 4 cluster

Tabel 9. Hasil Nilai DBI

Cluster	DBI
2	1,2759
3	0,9937
4	0,8737

4. KESIMPULAN

Berdasarkan hasil analisis maka dapat disimpulkan bahwa:

1. Penerapan Algoritma *Fuzzy C-Means* berdasarkan tingkat pengangguran dan tingkat kemiskinan untuk mengelompokkan wilayah kabupaten/kota di Jawa Timur, diperoleh cluster optimum dengan hasil 4 cluster berdasarkan nilai *davies bouldin index* 0,8737.
2. Hasil pengelompokan dengan 4 cluster pada :
 - a. C1 terdapat 15 wilayah terdiri dari Kab. Ponorogo, Kab. Tulungagung, Kab. Blitar, Kab. Kediri, Kab. Malang, Kab. Lumajang, Kab. Jember, Kab. Banyuwangi, Kab. Pasuruan, Kab. Mojokerto, Kab. Jombang, Kab. Nganjuk, Kab. Madiun, Kab. Magetan, Kota Probolinggo
 - b. C2 terdapat 4 wilayah terdiri dari Kab. Probolinggo, Kab. Bangkalan, Kab. Sampang, Kab. Sumenep
 - c. C3 terdapat 10 wilayah terdiri dari Kab. Pacitan, Kab. Trenggalek, Kab. Bondowoso, Kab. Situbondo, Kab. Ngawi, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Gresik dan
 - d. C4 terdapat 9 wilayah terdiri dari Kab. Sidoarjo, Kab. Kediri, Kab. Blitar, Kota Malang, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu.

REFERENSI

- [1] Badan Pusat Statistik Provinsi Jawa Timur, “Tingkat Pengangguran Terbuka (TPT) Menurut Kabupaten/Kota, 2001 - Agustus 2018.” [Online]. Available: <https://jatim.bps.go.id/dynamictable/2018/11/05/432/tingkat-pengangguran-terbuka-tpt-menurut-kabupaten-kota-2001---agustus-2018.html>.
- [2] Badan Pusat Statistik Provinsi Jawa Timur, “Persentase Penduduk Miskin Menurut Kabupaten/Kota di Provinsi Jawa Timur Tahun 2012-2017.” [Online]. Available: <https://jatim.bps.go.id/statictable/2018/01/30/755/persentase-penduduk-miskin-menurut-kabupaten-kota-di-provinsi-jawa-timur-tahun-2012-2017.html>.
- [3] Badan Pusat Statistik Provinsi Jawa Timur, “Laporan Eksekutif Keadaan Angkatan Kerja di Jawa Timur Tahun 2011-2012.”
- [4] N. Komariyah and M. S. Akbar, “Pengelompokan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan Indikator Kemiskinan Dengan Metode Cluster Analisis.”
- [5] M. Widiastuti and E. Yusuf, “Pemetaan Kemiskinan Kabupaten/Kota di Provinsi Jawa Tengah Tahun 2002 dan 2010 Menggunakan Analisis Klaster,” *Diponegoro J. Econ.*, vol. 1, no. 1, pp. 1–14, 2012.
- [6] F. Ramdhani *et al.*, “Pengelompokan Provinsi di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster,” vol. 4, pp. 875–884, 2015.
- [7] Sundar, M. Chitradevi, and G. Geetharamani, “An Analysis On The Performance Of K-Means Clustering Algorithm For Cardiogram Data Clustering,” *Int. J. Comput. Sci. Appl.*, vol. 2, no. 5, 2012.