

Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Menggunakan Orange Data Mining (Studi Kasus: Kabupaten/Kota Provinsi Sulawesi Selatan)

Muh. Nurfaizi Mursidin¹, Adinda Ridhallah Azhan², Komang Aryasa*³,
Risnayanti Andi Djamro⁴

^{1,2}Universitas Dipa; Jl.Perintis Kemerdekaan KM.09, 0411-587194/0411-588283

³Jurusan Teknik Informatika, Universitas Dipa Makassar

e-mail :¹muhnurfaizi2@gmail.com, ²adindaridhallah@gmail.com

Abstrak

Menurut data dari Badan Pusat Statistika (BPS) berdasarkan Anggaran Pendapatan dan Belanja Daerah (APBD), provinsi Sulawesi Selatan termasuk dalam 10 provinsi terkaya di Indonesia. Penduduk miskin di provinsi Sulawesi Selatan pada Maret 2021 berjumlah 786.980 menurun hingga 765.460 penduduk miskin pada September 2021. Penduduk yang berada di pedesaan mendominasi turunnya angka kemiskinan di provinsi Sulawesi Selatan. Berbanding terbalik dengan penduduk di perkotaan, presentase pada maret 2021 sebesar 4,77 persen naik hingga 4,89 persen di bulan September 2021. Untuk mengatasinya, diperlukan penanganan yang berbeda sesuai kondisi setiap Kabupaten/Kota, dengan salah satu cara yaitu mengelompokkan karakteristik setiap wilayah berdasarkan indikator kemiskinan. Pada penelitian ini, clustering dilakukan dengan metode algoritma K-Means menggunakan Orange Data Mining berdasarkan indikator kemiskinan yaitu persentase penduduk miskin usia 15 tahun ke atas, angka melek huruf dan angka partisipasi sekolah, persentase tidak bekerja, bekerja di kegiatan informal, bekerja di kegiatan formal, persentase bekerja di sektor pertanian dan bekerja bukan di sektor pertanian, persentase pengeluaran perkapita untuk makanan, persentase rumah tangga miskin yang menggunakan air layak, persentase rumah tangga miskin yang menggunakan jamban sendiri/bersama, serta jumlah beras yang diterima perbulan. Hasil penelitian didapatkan 6 cluster berdasarkan karakteristik daerah dari nilai rata – rata tertinggi dan terendah dari setiap indikator kemiskinan.

Kata Kunci : Kemiskinan, Cluster, K-Means, Orange Data Mining

Abstrack

According to data from the Central Statistics Agency (BPS) based on the Regional Budget (APBD), South Sulawesi Province is included in the 10 richest provinces in Indonesia. Poor population in South Sulawesi Province in March 2021 totaling 786,980 decreased to 765,460 poor people in September 2021. Residents in the countryside dominated the decline in poverty in South Sulawesi Province. Inversely proportional to the population in urban areas, the percentage in March 2021 amounted to 4.77 percent up to 4.89 percent in September 2021. To overcome this, different handling is needed according to the conditions of each district/city, with one way that is grouping the characteristics of each region based on poverty indicators. In this study, clustering was carried out by the K-Means algorithm method using Orange Data Mining based on poverty indicators, namely the percentage of poor people aged 15 years and over, literacy rates and school participation rates, the percentage of not working, working in

informal activities, working in formal activities, Percentage of Work in the Agriculture Sector and Working Not in the Agriculture Sector, Percentage of Per capita expenditure for food, the percentage of poor households that use decent water, the percentage of poor households that use their own/joint latrines, as well as the amount of rice received per month. The results obtained 6 clusters based on the characteristics of the area of the highest and lowest average value of each poverty indicator.

Keywords: *Poverty, Cluster, K-Means, Orange Data Mining*

1. PENDAHULUAN

Kemiskinan merupakan salah satu masalah Indonesia yang belum ditemukan secercah harapan atau solusi yang benar-benar efektif untuk mengatasinya. Kemiskinan sering diremehkan, tetapi kemiskinan adalah gejala yang kompleks. Padahal, pendapatan rendah yang sering menjadi indikator kemiskinan merupakan salah satu mata rantai yang menciptakan mata rantai kemiskinan. Pendapatan per kapita juga memiliki dampak yang signifikan terhadap pembangunan daerah. Menurut data dari Badan Pusat Statistika (BPS) berdasarkan Anggaran Pendapatan dan Belanja Daerah (APBD), provinsi Sulawesi Selatan termasuk dalam 10 provinsi terkaya di Indonesia. Penduduk miskin di provinsi Sulawesi Selatan pada Maret 2021 berjumlah 786.980 menurun hingga 765.460 penduduk miskin pada September 2021. Penduduk yang berada di pedesaan mendominasi turunnya angka kemiskinan di provinsi Sulawesi Selatan. Berbanding terbalik dengan penduduk di perkotaan, presentase pada maret 2021 sebesar 4,77 persen naik hingga 4,89 persen di bulan September 2021[1].

Untuk mengatasinya, diperlukan perlakuan yang berbeda antar daerah tergantung situasi di masing – masing kabupaten/kota di provinsi Sulawesi Selatan. Pengelompokan karakteristik wilayah dapat dilihat dari persentase penduduk miskin usia 15 tahun ke atas, angka melek huruf dan angka partisipasi sekolah, persentase tidak bekerja, bekerja di kegiatan informal, bekerja di kegiatan formal, persentase bekerja di sektor pertanian dan bekerja bukan di sektor pertanian, persentase pengeluaran per kapita untuk makanan, persentase rumah tangga miskin yang menggunakan air layak, persentase rumah tangga miskin yang menggunakan jamban sendiri/bersama, serta jumlah beras yang diterima perbulan sesuai dengan indikator kemiskinan yang telah ditentukan publikasi data kemiskinan Kabupaten/Kota yang disajikan oleh BPS.

2. METODE PENELITIAN

A. Clustering

Clustering adalah metode pengelompokan instance (sample) ke dalam beberapa group atau subset atau *cluster* berdasarkan “kemiripannya” dengan instance lain. Pada prinsipnya *clustering* hampir sama dengan *classification*, hanya saja dataset yang digunakan tidak berpasangan atau tidak berlabel (non labeled). Dataset semacam ini dapat dijumpai di sekitar kita dan jenisnya relatif cukup banyak. Sebagai contoh, perhatikan dataset jenis buku dan jumlah halamannya[2].

B. Algoritma K-Means

K-Means adalah algoritma yang umum digunakan untuk *clustering* data. Prinsip utama dari *K-Means* adalah menyusun k prototipe atau centroid dari sekumpulan data n-dimensi.

Algoritma *K-Means* menelompokkan objek atau data yang ada ke dalam *k* atau *cluster*. Untuk melakukan *clustering*, terlebih dahulu kita perlu mengetahui nilai *k*. Algoritma *K-Means* menggunakan ukuran ketidakmiripan untuk mengelompokkan objek. Ketidakmiripan ini didefinisikan oleh konsep jarak. dua objek dikatakan mirip jika jarak antara keduanya cukup kecil. Semakin dekat, maka semakin tinggi kemiripannya. Semakin tinggi nilai jaraknya, maka semakin tinggi pula ketidakmiripannya.

Proses algoritma *K-Means* [2] :

1. Tentukan *k* sebagai jumlah cluster yang akan dibentuk dan tentukan centroid awal secara random

2. Hitung jarak setiap data ke centroid cluster menggunakan persamaan Euclidean

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \dots\dots\dots (1)$$

3. Kumpulkan data ke dalam cluster dengan jarak terdekat menggunakan persamaan

$$\text{Min } \sum_{i=1}^k d_{ik} = \sqrt{\sum_n^m (C_{ij} - C_{kj})^2} \dots\dots\dots (2)$$

4. Hitung pusat cluster yang baru menggunakan persamaan

$$C_{ij} = \frac{\sum_{l=1}^p x_{lj}}{p} \dots\dots\dots (3)$$

Dimana : C_{ij} = cluster ke-*k*

5. Ulangi langkah 2 sampai 4 hingga sudah tidak ada lagi data yang berpindah ke cluster yang lain.

C. Metode Davies-Bouldin Index (DBI)

David L. Davies dan Donald W. Bouldin memperkenalkan sebuah metode bernama *Davies-Bouldin Index* (DBI) yang digunakan untuk menilai *cluster*. penilaian menggunakan *Davies-Bouldin Index* memiliki skema evaluasi internal *cluster*, dimana baik atau tidaknya hasil *cluster* dilihat dari jumlah dan kedekatan antar data hasil *cluster*. *Davies-Bouldin Index* adalah salah satu cara untuk mengukur efektivitas *cluster* dalam teknik *clustering*. Kohesi diartikan sebagai jumlah dari kedekatan data dengan pusat *cluster* dari *cluster* yang dilacak. Sedangkan separasi didasarkan pada jarak antar *cluster* dari pusatnya. Menggunakan *Davies-Bouldin Index* sebagai pengukuran, memaksimalkan jarak *inter-cluster* antara *cluster* C_i dan C_j dan pada waktu yang sama mencoba meminimalkan jarak antara titik dalam *cluster*. ketika jarak antar-*cluster* maksimum, berarti kesamaan karakteristik antara setiap *cluster* kecil, dan perbedaan antar-*cluster* terlihat lebih jelas. ketika jarak intra-*cluster* minimal berarti antara objek dalam *cluster* memiliki tingkat kesamaan karakteristik yang tinggi[3]. Untuk menghitung *Davies-Bouldin Index* memiliki beberapa tahapan yaitu:

1. Sum of Square Within-cluster (SSW) untuk mengetahui kohesi ke-*i* menggunakan persamaan sebagai berikut :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \dots\dots\dots (3)$$

2. Sum of Square Between-cluster (SSB) untuk mengetahui separasi antar cluster menggunakan persamaan sebagai berikut :

$$SSB_{i,j} = d(c_i, c_j) \dots\dots\dots (4)$$

3. Ratio (Rasio) untuk mengetahui nilai perbandingan antar cluster ke-*i* dan cluster ke-*j* menggunakan persamaan sebagai berikut :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \dots\dots\dots (5)$$

4. Davies-Bouldin Index untuk mencari nilainya digunakan nilai yang diperoleh dari persamaan (5) dengan menggunakan persamaan sebagai berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{l \neq j} (R_{i,j}) \dots\dots\dots (6)$$

Dari rumus tersebut, k adalah jumlah *cluster*. Semakin kecil nilai *Davies-Bouldin Index* (DBI) yang diperoleh (non-negatif ≥ 0), semakin baik pula *cluster* yang diperoleh dari *clustering* menggunakan algoritma *clustering*.

3. HASIL DAN PEMBAHASAN

A. Analisa Data Kemiskinan

Pada penelitian ini, kami mengambil data kemiskinan pada tahun 2021 yang bersumber dari website BPS. Dimana kami memilih beberapa data yang kami rasa dapat menjadi sumber atau acuan dalam penelitian kali ini dalam melakukan *clustering* data kemiskinan. Adapun dataset sebagai berikut :

Tabel 1 Tabel indikator kemiskinan

KOTA/KABUPATEN	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
Kepulauan Selayar	33,7	47,5	18,8	100,0	96,1	100,0	92,8	36,3	47,7	16,0	34,5	29,3	71,9	82,3	58,2
Bulukumba	31,2	53,8	15,1	96,5	91,6	100,0	69,1	51,0	34,3	14,7	39,2	9,8	70,1	78,8	95,0
Bantaeng	57,2	30,1	12,6	100,0	84,2	100,0	67,1	35,9	56,4	7,7	44,3	19,8	71,3	85,9	78,8
Jeneponto	43,3	40,2	16,5	100,0	81,3	99,8	78,1	39,4	53,6	7,0	45,9	14,8	69,7	95,1	76,6
Takalar	28,8	51,9	19,3	100,0	91,9	94,2	81,4	49,3	32,4	18,3	23,6	27,1	65,6	94,5	91,1
Gowa	34,4	53,5	12,1	98,8	91,3	100,0	85,9	30,8	63,7	5,5	56,8	12,4	71,5	87,6	100,0
Sinjai	37,1	46,5	16,4	100,0	93,8	100,0	87,9	49,2	44,1	6,7	42,2	8,5	66,7	52,7	95,9
Maros	29,5	58,4	12,1	97,7	92,7	100,0	99,7	44,0	39,5	16,6	35,3	20,8	68,9	73,5	76,6
Pangkajene	16,7	64,3	19,0	100,0	97,4	100,0	93,8	50,6	30,3	19,1	26,7	22,7	65,2	75,1	90,5
Barru	27,8	50,9	21,3	96,7	94,8	94,2	100,0	52,9	32,7	14,4	26,3	20,8	66,4	92,5	94,2
Bone	38,9	47,6	13,6	97,3	88,1	96,4	85,5	53,1	41,1	5,9	37,6	9,3	63,0	84,5	93,5
Soppeng	33,0	49,6	17,4	100,0	99,6	94,0	79,6	58,7	33,5	7,8	32,8	8,5	60,4	88,9	100,0
Wajo	29,3	51,4	19,3	100,0	31,6	100,0	95,7	61,7	29,1	9,2	22,4	15,9	66,6	80,7	100,0
Sedenreng Rappang	34,5	48,6	16,9	100,0	96,2	97,4	100,0	43,0	36,6	20,4	32,3	24,7	63,6	97,1	88,0
Pinrang	43,2	45,3	11,5	100,0	91,8	99,7	84,6	51,9	35,3	12,9	26,9	21,2	64,1	83,7	92,9
Enrekang	22,3	54,6	23,1	100,0	98,7	100,0	100,0	48,7	46,9	4,4	44,5	6,8	69,1	74,4	95,2
Luwu	21,7	53,4	24,9	100,0	96,4	100,0	89,8	50,2	34,9	14,9	35,3	14,5	65,5	84,2	82,8
Tana Toraja	29,0	47,6	23,4	97,4	86,6	100,0	81,6	27,4	67,9	4,7	63,7	8,9	73,5	63,6	86,9
Luwu Utara	17,4	52,3	30,3	100,0	97,6	100,0	100,0	48,9	36,7	14,4	34,5	16,6	69,2	88,5	90,6
Luwu Timur	27,8	43,6	28,7	100,0	92,9	96,2	93,2	54,9	30,4	14,7	39,2	8,9	64,4	73,3	86,0
Toraja Utara	26,3	41,6	32,1	100,0	93,9	100,0	97,4	38,1	46,1	15,8	46,0	15,9	73,3	60,9	92,9
Kota Makassar	16,3	45,0	38,8	100,0	98,1	100,0	79,8	56,0	26,2	17,8	1,1	43,0	57,0	100,0	100,0
Kota Pare-Pare	28,6	34,3	37,0	100,0	100,0	100,0	97,3	48,9	32,4	18,7	17,1	34,0	58,8	100,0	100,0
Kota Palopo	12,9	57,1	30,0	100,0	99,1	100,0	100,0	57,4	27,7	14,9	11,0	31,6	64,4	96,3	76,5

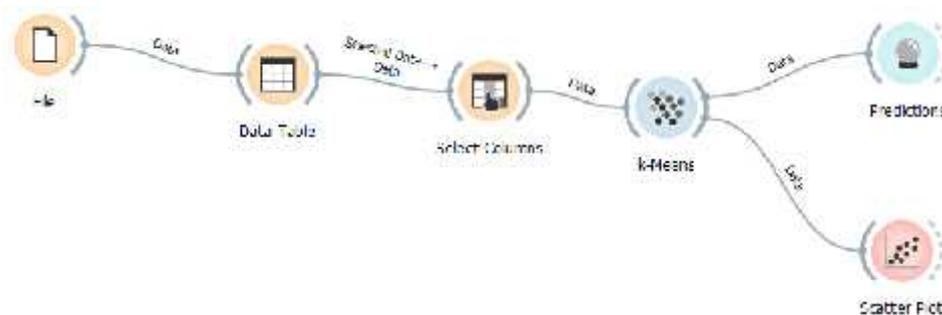
Dengan keterangan :

Tabel2 Keterangan indikator kemiskinan

x1	<SD	x9	Bekerja di kegiatan informal
x2	Tamat SD/SMP	x10	Bekerja di kegiatan formal
x3	>SMA	x11	Bekerja di sektor pertanian
x4	Angka melek huruf 15-24	x12	Bekerja bukan di sektor pertanian
x5	Angka melek huruf 15-55	x13	Persentase Pengeluaran Perkapita untuk Makanan
x6	Angka partisipasi sekolah 7-12	x14	Persentase Rumah Tangga Miskin yang Menggunakan Air Layak
x7	Angka partisipasi sekolah 13-15	x15	Persentase Rumah Tangga Miskin yang Menggunakan Jamban sendiri
x8	Tidak bekerja	x16	Persentase penduduk miskin

B. Skenario Penelitian

Menampilkan design widget text clustering dengan menggunakan Orange Data Mining seperti gambar dibawah ini :

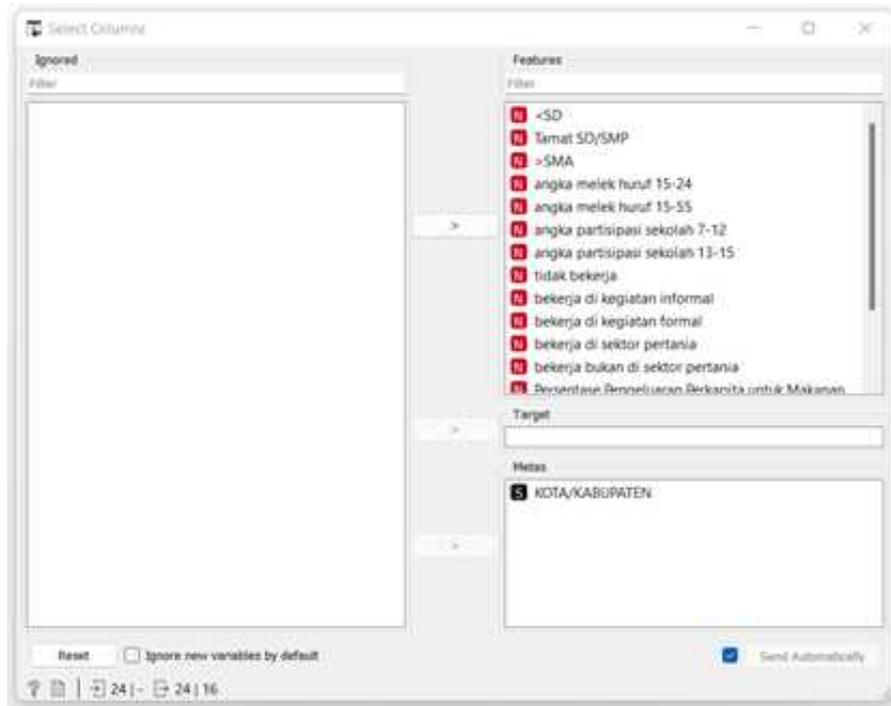


Gambar 1 Skenario penelitian menggunakan Orange Data Mining

Data yang telah diperoleh dan telah diseleksi akan di inputkan dalam widget file yang setelah itu dilakukan pengecekan pada data table untuk melihat apa data yang digunakan telah tepat, setelah itu makan widget data table di hubungkan ke widget k-Means untuk dilakukannya proses clustering, setelah itu di hubungkan ke widget predictions untuk melihat cluster yang telah terbentuk.

C. Widget File

Widget File merupakan widget yang berfungsi untuk memuat data yang ingin digunakan untuk mlakukan clustering, di dalam widget dapat kita lihat indikator dan atribut yang kita gunaka. Dapat dilihat sebagai berikut :



Gambar 4 Menu select collumn

F. Processing K-Means

Dalam widget ini kita mengatur memilih beberapa cluster yang ingin kita buat, serta berapa kali kita akan melakukan iterasi.



Gambar 5 Menu widget K-Means Orange Data Mining

G. Prediction

Setelah melakukan proses cluster pada widget K-Means tadi, untuk hasilnya menggunakan widget prediction. Didalam widget tersebut akan menampilkan data sebagai berikut :

KOTA/KABUPATEN	Cluster	Silhouette	+SD	Tingkat SCS/SMP	+SMA	ka kelas huruf 11	ka kelas huruf 11	partisipasi siswa	partisipasi siswa	tidak bekerja	ja di kegiatan ek	ja di kegiatan fo	ja di kegiatan
Kepulauan Selar	C1	0.517141	13.69	47.49	18.82	100.00	96.14	100.00	92.76	36.28	47.71	16.01	34.46
Takalar	C1	0.606721	28.83	51.87	33.30	100.00	91.91	94.18	81.37	48.26	32.40	18.34	27.62
Maros	C1	0.336793	29.54	58.23	32.11	87.67	92.68	100.00	89.72	43.97	38.45	35.58	33.28
Pangkajene	C1	0.591615	16.72	44.32	18.96	100.00	97.38	100.00	93.77	50.58	30.28	18.14	28.70
Barru	C1	0.416369	27.80	50.90	21.30	96.69	94.64	94.24	100.00	52.52	32.71	14.37	26.30
Soppeng	C1	0.537131	32.99	49.64	17.17	100.00	99.59	93.96	79.60	56.74	33.46	7.80	32.80
Sedereng Rappang	C1	0.589133	14.51	48.59	16.90	100.00	96.16	97.38	100.00	43.01	36.63	20.56	32.29
Pinrang	C1	0.54186	43.23	45.25	11.52	100.00	91.77	99.72	84.84	51.91	35.25	12.85	24.86
Luwu	C1	0.581411	21.74	53.37	24.90	100.00	96.36	100.00	88.83	50.22	34.89	34.89	35.28
Luwu Utara	C1	0.580981	17.43	52.25	30.32	100.00	97.57	100.00	100.00	48.80	36.70	18.41	34.48
Luwu Timur	C1	0.510057	27.75	43.55	26.70	100.00	93.91	96.11	93.55	54.93	36.42	14.63	39.20
Kota Makassar	C1	0.609038	16.20	44.96	38.75	100.00	98.06	100.00	78.84	55.97	26.23	17.60	3.06
Kota Pare-Pare	C1	0.603736	28.63	34.34	37.03	100.00	100.00	100.00	87.28	46.91	32.37	18.72	17.07
Kota Palopo	C1	0.620787	12.92	57.05	30.02	100.00	99.09	100.00	100.00	37.42	27.72	34.86	32.85
Wajo	C2	65	29.29	51.40	18.21	100.00	113.7	100.00	95.68	61.70	28.13	3.50	22.38
Bulukumba	C3	0.497446	11.15	51.75	15.09	96.46	91.64	100.00	85.52	51.02	14.31	14.66	39.16
Bantaeng	C3	0.548936	17.24	30.14	12.62	100.00	84.17	100.00	67.05	15.91	16.41	7.68	44.28
Jeneponto	C3	0.558707	43.29	40.17	16.54	100.00	81.29	88.81	78.12	39.36	33.63	7.01	45.87
Goa	C3	0.567908	34.41	53.45	12.14	96.78	91.25	100.00	85.91	30.80	43.69	15.1	56.79
Sinjai	C3	0.562187	17.10	46.54	18.23	100.00	93.77	100.00	87.88	49.23	44.11	6.66	42.34
Dona	C3	0.500017	38.85	47.61	13.25	87.25	88.14	96.43	85.43	53.08	41.05	3.88	37.64
Eurekang	C3	0.511871	22.32	54.61	23.07	100.00	98.72	100.00	100.00	37.42	27.72	34.86	32.85
Tana Toraja	C3	0.602807	28.99	47.61	23.39	97.39	86.37	100.00	81.63	27.40	67.90	4.70	63.67
Toraja Utara	C3	0.581689	26.29	41.57	32.14	100.00	93.86	100.00	97.38	38.13	46.07	13.81	46

Gambar 6 Tabel hasil clustering Orange Data Mining

H. Pengujian Menggunakan Metode Davies-Bouldin Index (DBI)

- Menghitung Sum of Square Within-cluster (SSW) pada pengujian sebelumnya pengelompokkan menggunakan 3 cluster terdiri dari cluster 1, cluster 2, cluster 3 dengan data berikut :

Tabel 3 Hasil Clustering menggunakan Orange Data Mining

Kota/Kabupaten	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	cluster
Kepulauan Selayar	33.69	47.49	18.82	100	96.14	100	92.76	36.28	47.71	16.01	34.46	29.26	71.86	82.27	38.22	12.43	C1
Takalar	28.83	51.87	19.3	100	91.91	94.18	81.37	48.26	32.4	18.34	23.62	27.12	62.30	94.49	91.07	8.23	C1
Maros	29.54	58.23	12.11	97.67	92.68	100	99.72	43.97	39.45	16.58	35.28	20.75	68.94	73.52	76.64	9.57	C1
Pangkajene	16.72	44.32	18.96	100	97.38	100	93.77	50.58	30.28	19.14	20.7	22.71	85.23	75.07	90.54	14.28	C1
Barru	27.8	50.9	21.3	96.69	94.84	94.24	100	52.92	32.71	14.37	26.3	20.78	66.35	92.47	94.21	8.88	C1
Soppeng	32.99	49.64	17.17	100	99.59	93.96	79.6	58.74	33.46	7.8	32.8	8.47	60.44	88.94	100	7.53	C1
Sedereng Rappang	14.51	48.59	16.9	100	96.16	97.38	100	43.01	36.63	20.56	32.29	24.71	63.56	97.08	87.99	5.04	C1
Pinrang	43.23	45.25	11.52	100	91.77	99.72	84.84	51.91	35.25	12.85	26.86	21.24	64.14	83.74	92.89	8.81	C1
Luwu	21.74	53.37	24.90	100	96.36	100	89.83	50.22	34.89	14.89	35.28	14.5	45.26	84.74	82.76	13.53	C1
Luwu Utara	17.43	52.25	30.32	100	97.57	100	100	48.80	36.7	14.41	34.48	16.63	69.22	88.46	90.59	13.59	C1
Luwu Timur	27.75	43.55	26.70	100	93.91	96.11	93.55	54.93	36.42	14.63	39.2	0.67	64.43	75.32	86.64	6.94	C1
Toraja Utara	26.29	41.57	32.14	100	93.86	100	97.38	38.13	46.07	13.81	46	15.87	73.25	60.88	92.94	11.99	C1
Kota Makassar	16.2	44.96	38.75	100	98.06	100	79.84	55.97	26.23	17.8	1.86	42.97	56.99	100	100	4.82	C1
Kota Pare-Pare	28.63	34.34	37.03	100	100	100	97.28	46.91	32.37	18.72	17.07	34.02	58.84	100	100	5.4	C1
Kota Palopo	12.92	57.05	30.02	100	99.09	100	100	37.42	27.72	14.86	10.95	31.63	64.43	96.26	6.49	8.14	C1
Wajo	29.29	51.4	19.31	100	113.7	100	95.68	61.7	29.13	9.16	22.38	15.91	66.56	80.71	100	6.46	C2
Bulukumba	11.15	51.75	15.09	96.46	91.64	100	89.12	51.02	34.31	14.66	39.16	9.82	70.07	78.83	95.01	7.43	C3
Bantaeng	17.24	30.14	12.62	100	84.17	100	67.05	35.91	56.41	7.68	44.28	19.81	71.26	85.87	78.78	9.41	C3
Jeneponto	43.29	40.17	16.54	100	81.29	88.81	78.12	39.36	33.63	7.01	45.87	14.77	69.72	95.06	76.37	14.28	C3
Goa	34.41	53.45	12.14	96.78	91.25	100	85.91	30.8	43.69	5.51	56.76	13.44	71.51	87.99	100	7.44	C3
Sinjai	17.1	46.54	18.23	100	93.77	100	87.88	49.23	44.11	6.66	42.34	8.53	66.72	52.88	95.89	8.84	C3
Dona	38.85	47.61	13.25	97.25	88.14	94.43	85.43	53.08	41.05	5.88	37.64	9.28	82.99	84.49	93.45	10.52	C3
Eurekang	22.32	54.61	23.07	100	98.72	100	100	48.71	46.92	4.38	44.49	6.8	69.1	74.38	95.2	12.47	C3
Tana Toraja	28.99	47.61	23.39	97.39	86.37	100	81.63	27.4	67.9	4.7	63.67	8.93	73.21	63.64	86.88	12.27	C3
Toraja Utara	26.29	41.57	32.14	100	93.86	100	97.38	38.13	46.07	13.81	46	15.87	73.25	60.88	92.94	11.99	C3

Selanjutnya menentukan SSW dengan cara yang pertama menghitung rata – rata dari indikator cluster

Tabel 4 Rata – rata dari setiap indikator

cluster	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16
c1	26,56	49,57	23,88	99,62	95,89	98,38	92,62	49,41	34,82	15,77	28,16	22,64	65,25	86,05	88,03	9,20
c2	29,29	51,40	19,31	100,00	31,57	100,00	95,68	61,70	29,13	9,16	22,38	15,91	66,56	80,71	100,00	6,46
c3	35,52	46,16	18,32	98,88	89,93	99,58	83,62	41,52	50,45	8,03	46,68	11,81	69,79	75,94	90,52	10,53

Setelah mengetahui rata – rata, kita menghitung jarak data dengan centroid yang diikuti

Tabel 5 Hasil perhitungan jarak data dengan centroid diikuti

jarak data dengan centeroid yang di iktuti
38,24620319
18,06338834
26,41905016
23,08809098
14,06379625
28,96194143
19,92475173
24,10121058
14,49186248
17,92277596
24,68644726
38,84951297
47,03833825
33,33943968
32,84327173
0
28,2786386
38,04057353
28,3817651
27,12233235
27,47458828
22,31692385
27,40626066
32,87865098

Setelah itu, menghitung nilai rata – rata jarak data dengan centroid yang diikuti dari setiap cluster yang telah terbentuk sebelumnya, dapat dilihat dibawah ini :

Tabel 6 Hasil perhitungan rata – rata data dengan centroid yang diikuti

jarak data dengan centeroid yang di iktuti	ssw
38,24620319	26,80267
18,06338834	SSW1
26,41905016	
23,08809098	
14,06379625	
28,96194143	
19,92475173	

Dapat dilihat nilai SSW 1 yaitu 26.80, SSW 2 29.02 dengan begitu nilai dari SSW dari setiap cluster yang terbentuk telah diketahui.

2. Menghitung Nilai Sum of Square Between-cluster (SSB)
Hasil dari perhitungan SSB yang didapatkan adalah sebagai berikut :

Tabel 7 Hasil SSB

SSB	HASIL
ssb12	68,31898
ssb23	72,32464
ssb13	34,63069

Hasil diatas didapatkan dengan menghitung jarak antar centroid yang telah didapatkan di awa untuk melakukan proses perhitungan SSW

3. Menghitung Ratio (Rasio)

Dibawah ini hasil dari perhitunga rasio :

Tabel 8 Hasil Rasio

r12	0,392317
r23	0,40127
r13	1,611991

Rasio diatas didapatkan dari perhitungan penjumlahan antara SSW dan selanjutnya dibagi dengan SSB.

4. Menghitung Davies-Bouldin Index (DBI)

Berikut adalah hasil dari perhitungan DBI :

Tabel 9 DBI dari K3

X	1	2	3	MAX	DBI
1	0	0,392317	1,611991	1,611991	1,208417
2	0,392317	0	0,40127	0,40127	
3	1,611991	0,40127	0	1,611991	

Setelah menghitung nilai rasio maka kita selanjutnya memasukkan nilai dari rasio tersebut kedalam table matrik, yang selanjutnya akan dicari nilai maksimal dari setiap matrik yang akan dihitung nilai rata – rata dari DBI, setelah proses perhitungan yang panjang maka didapatkan nilai DBI dari percobaan menggunakan K3 adalah 1,20842 yang dimana sudah cukup mendekati nilai nol.

I. Clustering menggunakan hasil Davies-Bouldin Index (DBI)

Berikut adalah hasil dari pengujian DBI :

Tabel 10 Hasil Pengujian DBI

Nilai K	DBI
K2	1,54
K3	1,21
K4	1,31
K5	1,19
K6	0,92

Dimana K6 merupakan nilai DBI terkecil dari semua proses pengujian yang telah dilakukan dengan hasil sebagai berikut :

Tabel 11 DBI K6

X	1	2	3	4	5	6	MAX	DBI
1	0	0,293585	0,916166	1,10523	0,539807	0,534269	1,10523	0,92
2	0,293585	0	0,266044	0,290037	0,260626	0,272476	0,293585	
3	0,916166	0,266044	0	1,098469	0,974586	0,591123	1,098469	
4	1,10523	0,290037	1,098469	0	1,125168	0,636554	1,125168	
5	0,539807	0,260626	0,974586	1,125168	0	0,789896	1,125168	
6	0,534269	0,272476	0,591123	0,636554	0,789896	0	0,789896	

Dapat dilihat pada tabel 10 hasil yang didapat melalui pengujian DBI dari setiap nilai K yang diatas diperoleh nilai terendah yaitu 0,92 yang dimana menggunakan K6, yang secara sederhananya jika nilai dari DBI semakin mendekati 0 maka disimpulkan clustering itu semakin baik atau maksimal. Selanjutnya clustering akan dilakukan menggunakan K6.

4. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan untuk mendapatkan clustering data kemiskinan kesimpulan yang didapatkan adalah :

1. Hasil Pada cluster 1 atau C1 dengan beranggotakan kepulauan selayar, takalar, maros, Pangkajene, Barru, Sidenreng Rappang, Luwu, Luwu Utara pada cluster ini memiliki permasalahan pada angka melek huruf dan partisipasi sekolah, sanitasi yang memiliki persentase yang rendah .Pada cluster 2 atau C2 dengan beranggotakan Wajo memiliki persentase yang tinggi pada sektor angka melek huruf dan partisipasi sekolah , Pendidikan Formal, Ketenagakerjaan, serta pada sektor sanitasi. Pada cluster 3 atau C3 dengan beranggotakan Bantaeng, Jeneponto, Gowa memiliki permasalahan pada Pendidikan Formal, Ketenagakerjaan, Pengeluaran perkapita, dikarenakan memiliki persentase yang cukup tinggi pada sektor tersebut. Pada cluster 4 atau C4 dengan beranggotakan Sinjai, Enrekang, Tana Toraja, Toraja Utara, dengan indikator tertinggi yaitu pada sektor angka melek huruf dan partisipasi sekolah, Ketenagakerjaan, penduduk miskin. Pada cluster 5 atau C5 dengan beranggotakan Bulukumba, Bone, Soppeng, Pinrang, Luwu Timur dengan indikator tertinggi yaitu pada sektor angka melek huruf dan partisipasi sekolah. Pada cluster 6 atau C6 dengan beranggotakan Kota Makassar, Kota Pare -Pare, Kota Palopo dengan indikator tertinggi masyarakatnya yang melanjutkan Pendidikan Formal, angka melek huruf dan partisipasi sekolah, Ketenagakerjaan, sanitasi
2. Berdasarkan hasil evaluasi *clustering* dengan *Davies-Bouldin Index* pada data set yang digunakan, mampu memperbaiki dalam peningkatan kualitas clustering dengan

menunjukkan hasil yang lebih baik dibandingkan penentuan nilai K-i secara konvensional. Dengan diketahuinya karakteristik di setiap wilayah, tentunya akan menjadi dasar yang kuat bagi penyelenggara pemerintahan dalam memberikan penanganan yang tepat dan cepat untuk mengatasi kemiskinan yang terjadi.

5. SARAN

Berdasarkan kesimpulan tersebut diatas, maka ada beberapa saran yang akan diutarakan pada penelitian kalini, yaitu:

1. Disarankan pada penelitian selanjutnya untuk menggunakan metode clustering yang lain terutama dalam kasus clustering kemiskinan ini dikarenakan hasil yang didapat kan dapat dan bisa berubah Ketika dilakukan clustering ulang.
2. Disarankan pada penelitian selanjutnya untuk menggunakan tools atau aplikasi lain dalam melakukan clustering data karena orange data mining kurang lengkap dalam fitur – fitur esensial seperti menguji performa dari clustering yang telah dibuat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada pembimbing dan penguji yang telah membimbing kami dalam pengerjaan skripsi sehingga jurnal ini bisa dibuat. Serta Penulis juga mengucapkan terima kasih kepada orang tua kami yang telah memberi dukungan moril maupun materiil.

DAFTAR PUSTAKA

- [1]. Triany, N. H., & Hasma, M.M. (2022). DATA DAN INFORMASI KEMISKINAN PROVINSI SULAWESI SELATAN 2021. Makassar : Badan Pusat Statistika Provinsi Sulawesi Selatan.
- [2]. Primartha, R. (2021). ALGORITMA MECHINE LEARNING. Bandung : Penertbit Infomatika.
- [3]. Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. Jurnal Sains Dan Manajemen, 9(1), 95–100. <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839>