

PENERAPAN ALGORITMA CLUSTERING UNTUK PENGELOMPOKAN TINGKAT KEMISKINAN PROVINSI BANTEN

Tb Ai Munandar

Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya
e-mail: tb.aimunandar@dsn.uharajaya.ac.id

Abstrak

Kemiskinan merupakan masalah sosial yang sudah tersebar di seluruh daerah di Indonesia, termasuk di Daerah Provinsi Banten. Meskipun Provinsi tersebut memiliki angka kemiskinan terendah, namun masalah ini adalah hal yang harus diutamakan. Jika menaik, maka akan menimbulkan dampak dari berbagai aspek, seperti tertutupnya akses pendidikan, tumbuhnya kriminalitas, bahkan bisa menghambat pertumbuhan ekonomi dikarenakan turunnya kualitas sumber daya manusia dan pendapatan yang dimiliki. Masalah ini perlu dicari solusi agar Pemerintah Banten mampu bertanggung jawab untuk mengontrol, bahkan mengurangi tingkat kemiskinan di Provinsi Banten. Penelitian ini dilakukan sebagai upaya untuk mengelompokkan wilayah di Provinsi Banten berdasarkan tingkat kemiskinan. Pendekatan yang dilakukan adalah algoritma Clustering, dimana ini merupakan metode Data Mining untuk mengelompokkan data menjadi beberapa grup berbeda berdasarkan kesamaan karakteristik setiap objeknya. data dibagi kedalam 3 cluster berdasarkan tingkat kemiskinan, metode clustering K-Means dan K-Medoids digunakan dalam penelitian ini, yang mempunyai perbedaan dari segi penentuan pusat cluster, hasil perhitungan menggunakan kedua metode tersebut lalu dibandingkan berdasarkan nilai Davies-Bouldin Index untuk menentukan hasil cluster yang lebih optimal. pendekatan ini juga didukung oleh penggunaan perangkat lunak R. Hasil Penelitian yang diharapkan mampu menjadi acuan kepada Pemerintah Daerah untuk pengambilan keputusan dalam menentukan arah pengentasan kemiskinan di masa mendatang. Hasil clustering menunjukkan pembagian 3 cluster sama rata pada K-medoid dan K-Means menggunakan RStudio, yaitu 3 Kabupaten/Kota dengan kemiskinan rendah, 3 Kabupaten/kota dengan kemiskinan sedang, dan 2 Kabupaten/Kota dengan kemiskinan tinggi. Meskipun demikian, nilai DBI pada algoritma K-Medoid lebih rendah dari K-Means dengan hasil masing-masing 0.582 dan 0.602

Kata kunci: Clustering, Data Mining, Davies-Bouldin Index, K-Medoid, K-Means, Tingkat Kemiskinan

I. PENDAHULUAN

Kemiskinan merupakan masalah global baik di negara maju maupun negara berkembang. Menurut [1], kemiskinan adalah ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar (makanan/non-makanan). Indonesia berada di peringkat ke-95 untuk kategori Produk Domestik Bruto (PDB) dengan angka \$13.998 [2]. Kemiskinan saat ini tidak hanya menjadi masalah nasional tetapi juga menjadi masalah di berbagai provinsi, salah satunya Provinsi Banten.

Berdasarkan data dari Badan Pusat Statistik dalam [3], tingkat kemiskinan di Provinsi Banten periode 2002-2020 cenderung fluktuatif, namun hanya beberapa periode yang mengalami penurunan angka kemiskinan. Pada September 2020 persentase penduduk miskin mengalami kenaikan jumlah penduduk sebanyak 857,64 ribu orang atau setara dengan 6,63 persen dikarenakan adanya pandemi Coronavirus Disease-19 (COVID-19) yang membatasi semua kegiatan penting sehingga berpengaruh pada kemiskinan.

Kenaikan angka kemiskinan juga terjadi pada tahun 2013 dan 2015 yang dikarenakan kenaikan harga barang pokok, yaitu sebanyak 5,89 persen pada September 2013 dan 5,90 persen pada Maret 2015.

Provinsi Banten saat ini telah melakukan banyak upaya untuk menurunkan tingkat kemiskinan diantaranya dengan membuat berbagai program dan kebijakan tentang pengentasan kemiskinan, baik bersifat makro maupun mikro, namun implementasi kebijakan di lapangan belum semua optimal dikarenakan keterbatasan dana untuk pelaksanaannya. Oleh karena itu, diperlukan pendekatan lain untuk pengentasan kemiskinan, salah satunya teknik Clustering.

Teknik Clustering merupakan metode dalam data mining yang mengelompokkan beberapa data ke dalam beberapa kumpulan berdasarkan karakter/objek yang sama [4]. Pada penelitian ini, hal yang dikelompokkan adalah Kabupaten/Kota di Provinsi Banten berdasarkan tingkat kemiskinan. Pengelompokan ini menggunakan kriteria terkait yang menentukan tingkat kemiskinan di Provinsi Banten,

yakni Jumlah Penduduk Miskin, Indeks Pembangunan Manusia (IPM), dan Produk Domestik Regional Bruto (PDRB) Atas Dasar Harga Konstan (ADHK).

Berdasarkan penelitian dari [5], Indeks Pembangunan Manusia (IPM) dan Produk Domestik Regional Bruto (PDRB) mempunyai pengaruh yang negatif namun sangat signifikan terhadap kemiskinan di Provinsi Banten, baik secara parsial maupun simultan. Menurut [6], pembangunan tidak hanya dilihat dari pendapatan ekonomis saja, tetapi juga harus meningkatkan kualitas sumber daya manusia. Pendidikan dan kesehatan yang baik akan meningkatkan produktivitas untuk meningkatkan pendapatan dalam PDRB per kapita. Pendapatan ini berpengaruh pada naiknya daya beli bahan pokok, sehingga ada juga kenaikan kesejahteraan dan pertumbuhan ekonomi daerah. Hal ini menyebabkan kemiskinan juga menurun.

Algoritma yang digunakan adalah K-Medoid dan K-Means. Keduanya mempunyai perbedaan, yakni penentuan pusat cluster. Metode K-Medoids menggunakan objek, sedangkan K-Means menggunakan rata-rata/means [7]. Meskipun berbeda, keduanya sama-sama merupakan metode partisional dan lebih cepat dibandingkan pendekatan hirarki [8]. Kedua metode ini berperan dalam membandingkan proses dan vektor pengelompokan yang diperoleh. Setelah memperoleh kedua hasil clustering, selanjutnya adalah memilih algoritma paling optimal menggunakan metode Davies-Bouldin Index (DBI). Hasil akhir DBI yang mendekati 0 dijadikan acuan dalam optimalisasi algoritma clustering, dikarenakan semakin kecil nilai yang diperoleh (non-negatif), maka semakin optimal [9].

Penelitian ini dapat menjadi alternatif untuk penyusunan kebijakan untuk pengentasan kemiskinan, karena bertujuan menunjukkan daerah mana yang menjadi prioritas Pemerintah Daerah untuk peningkatan sumber pembangunan, seperti lapangan pekerjaan dan sarana pendidikan, sehingga dapat mengontrol atau bahkan menurunkan angka kemiskinan.

II. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah penelitian terapan/applied research. Penelitian ini bertujuan menjadi acuan pemerintah daerah dalam peningkatan kinerja dalam program pengentasan kemiskinan untuk dilakukan secara merata di seluruh daerah di Provinsi Banten.

Tahapan pertama yaitu identifikasi terkait masalah tingginya tingkat kemiskinan di beberapa daerah di Provinsi Banten. Selain itu, studi pustaka dari beberapa sumber terkait, seperti literatur, artikel, jurnal, juga dilakukan.

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data-data terkait dengan kemiskinan di 8 Kabupaten/Kota di Provinsi Banten dalam rentang 11 tahun terakhir (2010-2020): 1) Jumlah Penduduk Miskin (per ribu jiwa); 2) Indeks Pembangunan Manusia (IPM) ;3) PDRB Atas

Dasar Harga Konstan (dalam juta Rupiah). Ketiga dataset ini berasal dari data yang dirilis oleh Badan Pusat Statistik (BPS) Provinsi Banten melalui situs web resmi <https://banten.bps.go.id/>.

2.2 Preparasi Data

Ketiga data asli kemudian ditransformasi menjadi tiga buah variabel menggunakan perhitungan rata-rata (perhatikan Tabel 1). Variabel yang digunakan mencakup rata-rata Indeks Pembangunan Manusia (2010-2020), rata-rata Jumlah Penduduk Miskin (2010-2020), dan rata-rata PDRB ADHK (2010-2020).

Tabel 1 Data Rata-Rata Kemiskinan

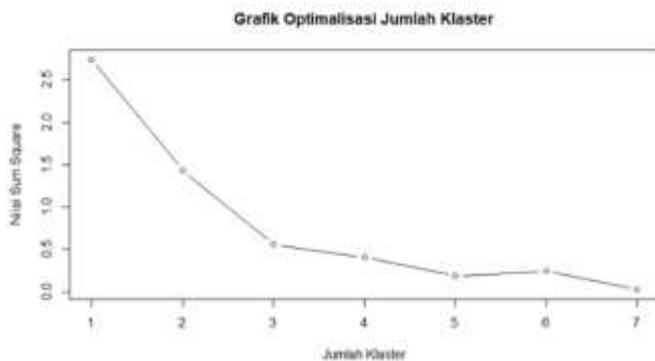
| Kab_Kota | Jumlah | IPM | PDRB_ADHK |
|----------------|--------|-------|---------------|
| Kab_Pandeglang | 117,91 | 62,46 | 16.118.430,18 |
| Kab_Lebak | 115,27 | 61,87 | 16.803.990,53 |
| Kab_Tangerang | 192,55 | 70,09 | 78.152.151,44 |
| Kab_Serang | 73,10 | 64,34 | 44.522.631,21 |
| Kota_Tangerang | 107,07 | 76,17 | 90.130.539,36 |
| Kota_Cilegon | 15,35 | 71,41 | 60.305.569,07 |
| Kota_Serang | 37,63 | 70,47 | 17.897.831,75 |
| Kota_Tangsel | 26,43 | 72,44 | 45.755.784,51 |

Sumber: BPS Provinsi Banten (data diolah)

Setelah data dikumpulkan, data tidak langsung dilakukan pengelompokan, dikarenakan terdapat perbedaan jarak yang terlalu jauh di ketiga variabel. Hal ini diperlukan normalisasi dengan rentang 0 sampai 1, terutama K-Means, jika ini tidak dilakukan, akan terdapat dominasi pada skala terbesar ketika perhitungan jarak [10].

2.3 Penentuan Jumlah Cluster

Selanjutnya adalah penentuan jumlah cluster. Dikarenakan jumlah sampel sedikit, akan digunakan data secara keseluruhan untuk melakukan pengelompokan. Jumlah cluster ditentukan hingga memperoleh hasil akhir, yaitu 3 cluster. Hal ini berdasarkan pendekatan Within Sum of Square (WSS) pada gambar di bawah ini.



Gambar 1. Visualisasi Pendekatan WSS

Gambar 1 memperlihatkan grafik jumlah cluster menurun drastis dari satu cluster hingga 7 cluster. Namun, jumlah cluster dilihat dari keseimbangan K dan nilai WSS, yang ditunjukkan dengan grafik menyiku atau knee [11].

2.4 K-Medoids

K-Medoids merupakan teknik pengelompokan yang menggunakan objek sebagai perwakilan (medoid) sebagai pusat cluster. K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada K-Means yang sensitif terhadap noise dan outlier. Hal ini menjadi penyebab objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Prosedur Algoritma K-medoid adalah sebagai berikut :

1. Tentukan pusat *cluster* sebanyak k (jumlah *cluster*).
2. Hitung setiap data/objek ke *cluster* terdekat menggunakan persamaan ukuran jarak *euclidean distance* dengan persamaan (1).

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

$d(x, y)$ = jarak antara data ke-x ke data k-y

i = dimensi data

x_i = koordinat data ke-xi ke dimensi i

y_i = koordinat data ke-yi ke dimensi i

3. Pilih secara acak objek pada masing-masing kelompok sebagai kandidat medoid baru (boleh random atau memilih dari data yang ada).
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru.
5. Hitung simpangan total *cost* ($S_{total\ jarak}$) dengan menghitung $S_{total\ jarak}$ = total jarak pada kandidat baru – total jarak pada kandidat lama. Jika $S_{total\ jarak} < 0$, maka kandidat *medoid* iterasi kedua menjadi *medoid* baru, sehingga harus dilakukan iterasi lagi

(lihat langkah 6), dan jika diperoleh $S_{total\ jarak} > 0$ maka iterasi berhenti.

6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoid* atau diperoleh $S_{total\ jarak} > 0$.

2.5 K-Means

K-Means adalah algoritma *clustering* yang mempartisi data dengan karakteristik sama ke dalam cluster/kelompok, sehingga data tersebut dikelompokkan ke dalam satu cluster sama dan data yang karakteristik berbeda masuk ke dalam cluster lain [12]. Setiap data kemudian ditetapkan pada setiap *cluster* berdasarkan pengamatan nilai rata-rata pada masing-masing *cluster*.

Tahap awal sama seperti K-Medoids, yaitu menentukan jumlah k dan menghitung jarak antar data dihitung menggunakan Euclidean (1). Namun, bedanya, centroid baru ditentukan berdasarkan dari nilai rata-rata/means dalam *cluster* yang didapatkan. Nilai rata-rata ini dihitung secara berulang pada proses awal [13].

Semua anggota kelompok/*cluster* (dari data asli) (x_q) dijumlahkan, lalu dibagi jumlah anggota *cluster* (ditunjukkan dengan N_k). Centroid baru ditunjukkan dalam μ_k , berikut adalah persamaannya (2).

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

2.6 Davies-Bouldin Index

Davies-Bouldin Index diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Perhitungan DBI berfungsi menguji hasil clustering dengan memaksimalkan jarak antar cluster. Suatu cluster memiliki skema clustering yang optimal jika nilai DBI minimal [14]. Semakin nilai DBI mendekati 0 (non-negatif), maka semakin baik [9].

Terdapat beberapa tahap yang harus dilakukan untuk menghitung DBI [15]. Tahap pertama adalah SSB (Sum of Square between Cluster), yang berfungsi mengetahui separasi atau jarak antar kelompok. Prosedurnya jarak antar centroid di hitung menggunakan Euclidean Distance (1) yang di tunjukkan sebagai $d(c_i, c_j)$ pada persamaan (3).

$$SSB_{ij} = d(c_i, c_j) \quad (3)$$

Tahap kedua adalah SSW (Sum of Square within Cluster). Caranya menghitung rata-rata dari jarak minimum terakhir yang diperoleh dari perhitungan Euclidean Distance. Untuk menghitung SSW digunakan persamaan (4).

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_i, c_i) \quad (4)$$

Keterangan:

m_i = jumlah data dalam *cluster* i

x = data dalam *cluster*

x_j = data pada *cluster* j

c_i = *centroid*/pusat *cluster* ke-i

$d(x_i, c_i)$ = jarak setiap data ke *centroid* i

Cluster yang baik adalah cluster yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Rasio bertujuan untuk mengetahui seberapa bagus perbandingan antara cluster i dengan cluster lainnya. Indeks i dan j merupakan presentasi jumlah cluster. Indeks n akan berlanjut tergantung sejumlah cluster yang dipilih dengan syarat ni tidak sama dengan n_i .

$$R_{ij,\dots,n} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

Keterangan:

SSW_i = hasil SSW untuk *cluster* i

SSB_{ij} = separasi dari *cluster* i dan *cluster* j

Nilai rasio yang diperoleh dari persamaan (5) digunakan untuk mencari nilai DBI dengan menggunakan persamaan ini (6). Semakin kecil nilai DBI, mendekati 0 selama tidak negatif, maka semakin baik hasil pengelompokan.

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

Keterangan:

K : *cluster* yang ada

$R_{ij,\dots,k}$: rasio antara cluster i dan j, i dan k, tergantung berapa *cluster*

Max $i \neq j$: dicari rasio antar *cluster* terbesar

III. HASIL DAN PEMBAHASAN

3.1 Pengelompokan Algoritma K-Medoids

Berdasarkan percobaan pada 2 iterasi, hasil tidak menunjukkan perubahan apapun terhadap vektor cluster, sehingga, dapat disimpulkan bahwa dari hasil pengklasteran dengan metode K-Medoid terdapat C1 sebanyak 3 Kabupaten/kota, C2 berjumlah 2 Kabupaten/Kota, dan C3 mencakup 3 Kabupaten/Kota. Tabel 3 menunjukkan pembagian anggota cluster yang berkorespondensi.

Tabel 3. Pembagian *Cluster* dengan K-Medoid

| Cluster | Kabupaten/Kota |
|---------|--|
| C1 | Kabupaten Pandeglang, Kabupaten Lebak, Kabupaten Serang |
| C2 | Kabupaten Tangerang, Kota Tangerang |
| C3 | Kota Serang, Kota Cilegon, Kota Tangerang Selatan |

3.2 Hasil Pengelompokan dengan K-Means

Hasil pengujian dengan K-Means juga tidak terdapat perbedaan dengan algoritma K-Medoid. Pengujian pun dilakukan dengan dua kali iterasi. Namun implementasi K-Means menggunakan R studio sangat bergantung pada penentuan *centroid*, sehingga seringkali mengalami pengacakan dan sulit mendapat hasil awal yang unik [16]. Maka dari itu dilakukan pengacakan (nstart) sebanyak 4 kali. Tabel 4 memperlihatkan hasil klaster dengan K-means

Tabel 4. Pembagian *Cluster* dengan K-Means (5)

| Cluster | Kabupaten/Kota |
|---------|--|
| C1 | Kabupaten Pandeglang, Kabupaten Lebak, Kabupaten Serang |
| C2 | Kabupaten Tangerang, Kota Tangerang |
| C3 | Kota Serang, Kota Cilegon, Kota Tangerang Selatan |

3.3 Hasil Evaluasi DBI

Setelah hasil vektor pengelompokan telah didapat, selanjutnya adalah menentukan algoritma paling optimal. Dalam kasus ini, penentuan optimasi algoritma clustering dilihat dari nilai akhir hasil Davies Bouldin Index (DBI). Kedua algoritma mempunyai kesamaan, yaitu nilai yang sama-sama mendekati 0 (non-negatif). Semakin kecil nilai DBI yang dimiliki, maka semakin optimal hasilnya.

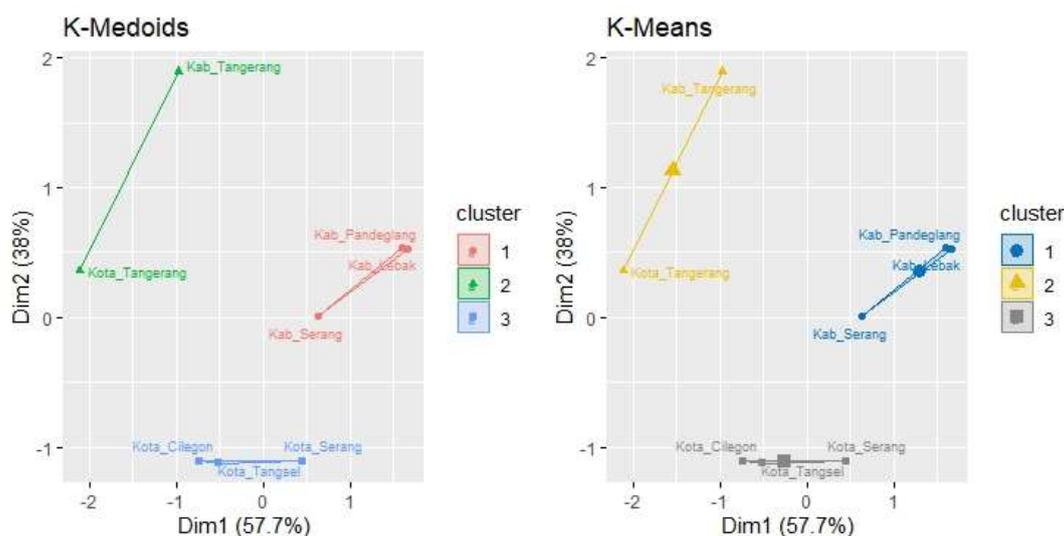
Tabel 5. Perbandingan Hasil DBI

| Algoritma Pengujian | Hasil DBI |
|---------------------|-----------|
| K-Medoid | 0.582 |
| K-Means | 0.602 |

Berdasarkan hasil di atas, dapat disimpulkan bahwa hasil dengan menggunakan K-Medoid jauh lebih optimal dibandingkan dengan K-Means. Sebagaimana dijelaskan dalam [17], K-Medoid mengatasi kelemahan K-Means yang sensitive pada pencilan/outlier, dan menggunakan perhitungan yang bergantung pada *centroid* dengan nilai cost terkecil.

3.4 Perbandingan Hasil Cluster

Gambar 2 merupakan visualisasi keluaran berupa grafik dari kedua pengelompokan yang diperoleh.



Gambar 2. Visualisasi Hasil Pengelompokan

Grafik di atas menghasilkan vektor pengelompokan yang dibagi sama rata, baik pada K-Medoids maupun pada K-Means. Vektor yang dimaksud terdapat pada Tabel 6.

Tabel 6 Hasil Akhir Pembagian Cluster

| Kab/Kota | Pembagian Cluster (K-Medoid) | Pembagian Cluster (K-Means) |
|----------------|------------------------------|-----------------------------|
| Kab_Pandeglang | C1 | C1 |
| Kab_Lebak | C1 | C1 |
| Kab_Tangerang | C2 | C2 |
| Kab_Serang | C1 | C1 |
| Kota_Tangerang | C2 | C2 |
| Kota_Cilegon | C3 | C3 |
| Kota_Serang | C3 | C3 |

Untuk memperoleh interpretasi vektor pengelompokan di atas, perlu dilakukan profilisasi pusat/centroid akhir, yaitu perhitungan rata-rata dari data asli pra-normalisasi. Dalam kasus ini, jika dua algoritma memperoleh vektor pengelompokan yang sama, maka rata-rata yang akan diperoleh juga sama. Hasil profilisasi yang dimiliki algoritma K-Medoid seperti diperlihatkan pada Tabel 7.

Tabel 7. Nilai Centroid Akhir

| Cluster | jumlah | ipm | Pdrb_adhk |
|---------|--------|-------|-------------|
| C1 | 102,10 | 62,89 | 25815017,31 |
| C2 | 149,81 | 73,13 | 84141345,40 |
| C3 | 26,47 | 71,44 | 41319728,44 |

Penduduk miskin akan berkurang jika dipicu oleh tingginya produktivitas/IPM dan juga PDRB/pendapatan yang

tinggi. Berikut adalah intepretasi yang didapat dari penelitian secara keseluruhan.

1. C3 merupakan kategori kelompok daerah dengan tingkat kemiskinan rendah, dengan dengan Jumlah Penduduk Miskin rendah, IPM sedang, dan PDRB ADHK sedang. Kota Serang, Kota Cilegon, Kota Tangerang Selatan termasuk dalam kategori ini
2. C1 dikategorikan sebagai kelompok daerah dengan tingkat kemiskinan sedang. Hal ini dikarenakan mempunyai jumlah penduduk miskin sedang, IPM rendah, dan PDRB ADHK rendah. Daerah dalam kategori ini adalah Kabupaten Pandeglang, Kabupaten Lebak, Kabupaten Serang merupakan.
3. C2 merupakan kelompok daerah dengan tingkat kemiskinan tinggi. Seluruh variabel yang bernilai tinggi di antara kategori lain. Daerah dalam kategori ini adalah Kabupaten Tangerang dan Kota Tangerang

IV. KESIMPULAN

Melalui proses analisis yang sudah dilakukan, penelitian ini memperoleh kesimpulan bahwa didapatkan 3 kelompok daerah terpilih berdasarkan tingkat kemiskinan. Namun, berdasarkan profilisasi centroid menggunakan rata-rata, "Kelompok dengan tingkat kemiskinan tinggi" didapat oleh Kabupaten Tangerang dan Kota Tangerang. "Kelompok dengan tingkat kemiskinan sedang" beranggotakan Kabupaten Pandeglang, Kabupaten Lebak, Kabupaten Serang. Sementara "Kelompok dengan tingkat kemiskinan rendah" adalah Kota Serang, Kota Cilegon, Kota Tangerang Selatan. Dengan demikian, kelompok yang harus diprioritaskan dalam pengentasan kemiskinan adalah kelompok berisi daerah dengan tingkat kemiskinan tertinggi.

Kedua algoritma yang digunakan dalam penelitian ini mempunyai proses dan vektor pembagian cluster terlihat sama.

Meskipun begitu, berdasarkan perolehan hasil DBI, didapat K-Medoid dengan 0.582, dan K-Means dengan nilai 0.602.

V. SARAN

Penelitian yang dilakukan hanya memvisualisasikan hasil kluster dalam bentuk single view. Setiap atribut terkluster berdasarkan hasil perhitungan jarak antara centroid dengan titik data. Visualisasi tidak merepresentasikan bagaimana setiap atribut membentuk kluster masing-masing dan tervisualisasi dalam bentuk multi-view. Teknik multi-view clustering ke depannya mungkin bisa digunakan sebagai salah satu alternative untuk mengelompokkan lebih dari satu indicator dan menghasilkan bentuk kluster multi-view.

REFERENSI

- [1] Badan Pusat Statistik Provinsi Banten, 2020, Provinsi Banten Dalam Angka 2020, Serang: BPS Provinsi Banten.
- [2] Ventura, L., 2020, Poorest Countries in the World 2020. Retrieved April 25, 2021, from Global Finance: <https://www.gfmag.com/global-data/economic-data/the-poorest-countries-in-the-world>
- [3] Sutrisna, 2021, Kemiskinan di Banten Kembali Naik di Atas Rata-rata Nasional, BPS Menyebut Karena Faktor Ini, Retrieved 31 December 2021, from kabarbanten.com: kabarbanten.pikiran-rakyat.com/seputar-banten/pr-591441129/kemiskinan-di-banten-kembali-naik-di-atas-rata-rata-nasional-bps-menyebut-karena-faktor-ini
- [4] Atmaja, E. H. S., 2019, Implementation of k-medoids clustering algorithm to cluster crime patterns in Yogyakarta, International Journal of Applied Sciences and Smart Technologies, 1(1), 33-44.
- [5] Syamsiah, 2019, Pengaruh Indeks Pembangunan Manusia dan Produk Domestik Regional Bruto Terhadap Tingkat Kemiskinan di Provinsi Banten Periode 2010-2016, Doctoral dissertation, UIN SMH BANTEN.
- [6] Prasetyoningrum, A. K., dan Sukmawati, U. S., 2018, Analisis Pengaruh Indeks Pembangunan Manusia (IPM), Pertumbuhan Ekonomi dan Pengangguran Terhadap Kemiskinan di Indonesia, Equilibrium: Jurnal Ekonomi Syariah, 6(2), 217-240.
- [7] Riyanto, B., 2019, Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan), KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer), 3(1). <https://doi.org/10.30865/komik.v3i1.1659>
- [8] Farachi, A. N, 2020, Perbandingan Algoritma K-Means dan Algoritma K-Medoids dalam Pengelompokan Komoditas Tanaman Biofarmaka di Provinsi Jawa Tengah Tahun 2018, Thesis, Universitas Islam Indonesia.
- [9] Dewi, D. A. I. C., dan Pramita, D. A. K., 2019, Analisis perbandingan metode elbow dan silhouette pada algoritma clustering K-Medoids dalam pengelompokan produksi kerajinan bali, Matrix: Jurnal Manajemen Teknologi dan Informatika, 9(3), 102-109.
- [10] Sutresno, S. A., Iriani, A., dan Sedyono, E., 2018, Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan, JuTISI (Jurnal Teknik Informatika dan Sistem Informasi), 4(3), 433-440
- [11] Kassambara, A., 2017, Practical Guide to Cluster Analysis in R Edition 1. STDHA.
- [12] Anggara, M., Sujaini, H., dan Nasution, H., 2016, Pemilihan distance measure Pada K-Means clustering Untuk Pengelompokan member di Alvaro fitness, JUSTIN (Jurnal Sistem dan Teknologi Informasi), 4(1), 186-191.
- [13] Kamila, I., Khairunnisa, U., dan Mustakim, M., 2019, Perbandingan Algoritma K-means dan K-medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau, Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi, 5(1), 119.
- [14] Religia, Y., dan Jaya, R. T. B., 2020, Pengelompokan Menggunakan Algoritma K-Medoid Untuk Evaluasi Performa Siswa, Jurnal Pelita Teknologi, Vol. 15 (1) 2020, pp.49-55
- [15] Dehotman, B.J., 2018, Peningkatan hasil evaluasi clustering davies-bouldin index dengan penentuan titik pusat cluster awal algoritma k-means, Thesis, Universitas Sumatera Utara, Medan
- [16] Hablum, R. J., Khairan, A., dan Rosihan, R., 2019, Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means, JIKO (Jurnal Informatika dan Komputer), 2(1), 26-33.
- [17] Munandar, T. A., 2019, Bahan Ajar Data Mining Dengan R Edisi Revisi 3, Fakultas Teknologi Informasi, Universitas Serang Raya, Serang.