

Pengelompokan Kabupaten Dan Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kesejahteraan Dengan Metode *K-Means* Dan *Density-Based Spatial Clustering Of Applications With Noise*

Maria Titah Jatipaningrum^{1)*}, Suci Eka Azhari²⁾, Kris Suryowati³⁾

^{1,2,3} Jurusan Statistika, IST Akprind Yogyakarta

Email: ¹titahjp@akprind.ac.id

²suciekaa29@gmail.com

³suryowati@akprind.ac.id

Abstrak

Provinsi Jawa Timur mempunyai kondisi kesejahteraan yang belum merata. Kondisi kesejahteraan yang belum merata tersebut ditandai dengan banyaknya penduduk miskin di Jawa Timur dan laju pertumbuhan ekonominya yang mengalami penurunan pada tahun 2020, hingga mencapai -2,39% karena dampak pandemi. Kesejahteraan dapat diukur melalui beberapa indikator, antara lain meliputi kepadatan penduduk, angkatan kerja, tingkat partisipasi angkatan kerja, dan tingkat pengangguran terbuka. Sebab demikian, untuk mengetahui pengelompokan kabupaten dan kota di Provinsi Jawa Timur berdasarkan tingkat kesejahteraan, maka dilakukan pengelompokan menggunakan metode *K-Means* dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Masing-masing dari kedua metode tersebut, dilakukan perhitungan jarak menggunakan jarak *Euclidean* dan *Manhattan*. Setiap jarak diuji validitasnya menggunakan *Davies-Bouldin Index* (DBI), *C-Index*, dan *Dunn Index*. Penelitian ini menghasilkan kesimpulan bahwa metode terbaik yaitu metode DBSCAN menggunakan jarak *Manhattan* dengan $MinPts = 2$ dan $eps = 4$ yang mempunyai nilai DBI terkecil sebesar 0,284 dan menghasilkan lebih sedikit *noise*, dengan 2 kluster terbentuk dan 5 *noise*. Kluster 1 terdiri dari 26 kabupaten, kluster 2 terdiri dari 7 kota, dan *noise* terdiri dari 5 kabupaten dan kota.

Kata Kunci: Kesejahteraan, K-Means, DBSCAN, Jarak Euclidean, Jarak Manhattan

Abstract

East Java Province has an uneven welfare condition. The uneven welfare conditions are indicated by a large number of poor people in East Java and the rate of economic growth which has decreased in 2020, reaching -2.39% due to the impact of the pandemic. Welfare can be measured through several indicators, while the indicators used to classify districts and cities in East Java among others include population density, labor force, labor force participation rate, and open unemployment rate. Thus, to find out the grouping of regencies and cities in East Java Province based on the level of welfare, grouping was carried out using the *K-Means* and *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) methods. For each of the two methods, distance calculations are performed using the *Euclidean* and *Manhattan* distances. Each distance was tested for validity using the *Davies-Bouldin Index* (DBI), *C-Index*, and *Dunn Index*. This study concludes that the best method is the DBSCAN method using *Manhattan* distance with $MinPts = 2$ and $eps = 4$ which has the smallest DBI value of 0.284, with 2 clusters formed and 5 *noise*. Cluster 1 consists of 26 regencies, cluster 2 consists of 7 cities, and *noise* consist of 5 regencies and cities.

Keywords: Welfare, K-Means, DBSCAN, Euclidean Distance, Manhattan Distance.

1. PENDAHULUAN

Indonesia merupakan salah satu negara yang mempunyai jumlah populasi manusia yang besar di dunia. Dilansir dari *website worldmeters.info* (2022), Indonesia menempati urutan keempat setelah China, India, dan Amerika Serikat sebagai negara dengan jumlah

populasi terbesar di dunia, dimana jumlah populasinya sekitar 278.467.636 jiwa. Jumlah ini terdiri dari masyarakat yang memiliki perbedaan kalangan, misalnya berdasarkan perbedaan pekerjaan, pendidikan, kesehatan, dan kemampuan daya beli. Perbedaan tersebut mengakibatkan tingkat kesejahteraan masyarakat di Indonesia berbeda-beda. Salah satu penyebab perbedaan kesejahteraan yaitu tingkat kemiskinan.

Jawa Timur merupakan salah satu provinsi di Indonesia dengan jumlah penduduk miskin yang terus meningkat dari Maret 2019 hingga Maret 2021, dimana pada Maret 2019 sebesar 4,1 juta jiwa yang meningkat pada Maret 2020 sebesar 4,4 juta jiwa dan Maret 2021 sebesar 4,58 juta jiwa (BPS, 2021). Selain itu dilansir dari *website* IDM (2021), provinsi Jawa Timur juga menempati peringkat ketiga setelah Bali dan DI Yogyakarta dalam Indeks Desa Membangun (IDM) pada tahun 2020 dengan skor 0,7217, dimana menunjukkan pembangunan yang tepat dari pemerintah sesuai dengan partisipasi masyarakat. IDM dibentuk berdasarkan tiga indeks, yaitu indeks ketahanan sosial, indeks ketahanan ekonomi, dan indeks ketahanan ekologi (IDM, 2021). Namun, apabila dilihat dari jumlah penduduk miskin yang terus meningkat maka hal ini tidak selaras dengan pembangunan yang tepat. Tentunya masih terdapat daerah yang ketahanan ekonominya masih rendah karena terus meningkatnya jumlah penduduk miskin.

Selain jumlah penduduk miskin, pertumbuhan ekonomi juga dapat mempengaruhi kesejahteraan. Dilansir dari *website* Badan Pusat Statistik Provinsi Jawa Timur (2021), laju pertumbuhan ekonomi di Provinsi Jawa Timur dari tahun 2017 hingga 2019 berkisar 5% dan mengalami kenaikan setiap tahunnya. Pada tahun 2019, laju pertumbuhan ekonomi Provinsi Jawa Timur mencapai 5,52%. Namun, pada tahun 2020 mengalami penurunan mencapai -2,39% dikarenakan dampak pandemi. Pandemi menyebabkan terbatasnya kegiatan produksi barang dan jasa, dimana hal ini menyebabkan laju pertumbuhan ekonomi di Provinsi Jawa Timur mencapai angka negatif. Tentunya, penurunan tersebut juga disebabkan karena tingkat kesejahteraan kabupaten dan kota di Provinsi Jawa Timur ada yang mengalami penurunan. Oleh karena itu, perlu adanya pengelompokan kabupaten dan kota di Provinsi Jawa Timur yang didasarkan pada 3 tingkat kesejahtraannya agar diperoleh kelompok-kelompok dengan tingkat kemiripan tertentu sehingga mempermudah dalam pembangunan sesuai dengan hasil pengelompokan.

Pengelompokan kabupaten dan kota khususnya di Provinsi Jawa Timur berdasarkan tingkat kesejahteraan perlu dilakukan untuk mengetahui pemerataan tingkat kesejahteraan yang sudah dilakukan di Provinsi Jawa Timur. Oleh karena itu, untuk mengetahui pengelompokan kabupaten dan kota di Provinsi Jawa Timur berdasarkan kesamaan karakteristik tingkat kesejahtraannya kedalam sejumlah kelompok yang heterogen maka akan digunakan metode *K-Means* dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Perbandingan kedua metode menggunakan dua metode jarak yaitu *Euclidean* dan *Manhattan*, yang dibandingkan berdasarkan nilai *Davies-Bouldin Index* (DBI), *Hubert & Levin C-Index*, dan *Dunn Index* untuk dipilih metode terbaik.

2. METODE

Penelitian ini menggunakan desain penelitian kuantitatif deskriptif yaitu penelitian yang menggunakan data berupa angka dan interpretasi hasil dilakukan dengan menganalisis dan menyusun data yang sudah ada sesuai dengan kebutuhan peneliti. Penelitian dilakukan di kabupaten dan kota yang berada di Provinsi Jawa Timur pada tahun 2020, yang dipublikasikan BPS Jawa Timur tahun 2021. Metode pengumpulan data yang dilakukan pada penelitian ini adalah menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur (<https://jatim.bps.go.id>) pada tahun 2020. Data yang digunakan merupakan data variabel-variabel pada tahun 2020.

A. Variabel

Variabel yang digunakan pada penelitian ini meliputi kepadatan penduduk (KP), jumlah angkatan kerja (AK), tingkat partisipasi angkatan kerja (TPAK), tingkat pengangguran terbuka (TPT), rata-rata lama sekolah (RLS), angka harapan hidup (AHH), pengeluaran per kapita (PP), dan persentase penduduk miskin (PPM):

B. Metode Analisis

Tahapan analisis data yang digunakan pada penelitian ini, meliputi:

- 1) Mengumpulkan data indikator kesejahteraan.
- 2) Melakukan analisis deskriptif.
- 3) Melakukan standarisasi data dan menentukan data *outlier*.
- 4) Melakukan uji asumsi pada analisis kluster, yaitu uji multikolinearitas.
- 5) Melakukan pengelompokan data menggunakan metode *K-Means*. Langkah-langkah analisis data menggunakan metode *K-Means clustering* sebagai berikut:

- a. Menentukan jumlah kluster optimal menggunakan *silhouette width* dengan rumus:

$$SI = \frac{1}{n} \sum_{i=1}^n s(i) \quad (1)$$

$$\text{dengan } s(i) = \frac{b(i)-a(i)}{\max[a(i),b(i)]}$$

Keterangan:

$a(i)$: Rata-rata jarak objek ke- i dengan objek lain dalam satu kluster

$b(i)$: Rata-rata jarak objek ke- i dengan objek lain pada kluster terdekat

- b. Menentukan titik pusat (*centroid*) kluster.

Menghitung jarak antar titik pada setiap objek dengan titik *centroid* menggunakan jarak *Euclidean* dan jarak *Manhattan*. Jarak *Euclidean* bertujuan mengukur jumlah kuadrat perbedaan nilai antar variabel (Kartikawati & Ambarwati, 2017). Rumus pengukuran jarak *Euclidean* sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (2)$$

Keterangan:

d_{ij} : Jarak antar objek ke- i dan pusat kluster ke- j

p : Jumlah variabel kluster

X_{ik} : Data dari objek ke- i pada variabel ke- k

X_{jk} : Pusat kluster ke- j pada variabel ke- k

Jarak *Manhattan* merupakan pengukuran jarak dengan menjumlahkan *absolute* dari variabel-variabel (Nugraheny, 2015). Adapun rumus dalam pengukuran jarak *Manhattan* adalah sebagai berikut:

$$d_{ij} = \sum_{k=1}^p |X_{ik} - X_{jk}| \quad (3)$$

Keterangan:

d_{ij} : Jarak antar objek ke- i dan pusat kluster ke- j

p : Jumlah variabel kluster

X_{ik} : Data dari objek ke- i pada variabel ke- k

X_{jk} : Pusat kluster ke- j pada variabel ke- k

- c. Mengelompokkan objek berdasarkan jarak minimum untuk menentukan anggota

klaster.

- d. Apabila terdapat anggota klaster yang berpindah maka kembali ke langkah b, hingga nilai *centroid* yang dihasilkan tetap dan anggota klaster tidak berpindah ke klaster lain.
 - e. Membuat plot klaster *K-Means* yang terbentuk.
- 6) Melakukan pengelompokan data menggunakan metode DBSCAN. Langkah-langkah analisis data menggunakan metode DBSCAN sebagai berikut:
- a. Menentukan parameter *MinPts* dan *epsilon*.
 - b. Menentukan nilai *p* atau titik awal secara acak.
 - c. Menghitung jarak antar semua titik terhadap *p* menggunakan jarak *Euclidean* dan jarak *Manhattan*.
 - d. Menentukan titik-titik yang *density-reachable* atau dapat menjangkau kepadatan terhadap *p* dan berada dalam radius *epsilon*. Apabila banyaknya titik yang *density-reachable* terhadap *p* lebih dari atau sama dengan *MinPts* maka titik *p* adalah *core point* sehingga klaster terbentuk dan dilanjutkan ke titik lain di sekitarnya yang merupakan *density-reachable* terhadap *p*. Apabila *p* merupakan *border point* (titik tepi) dan tidak ada titik yang *density-reachable* terhadap *p* maka dilanjutkan ke titik yang lain.
 - e. Melakukan pengulangan c dan d hingga semua titik telah diproses. Apabila terdapat titik yang tidak masuk ke dalam klaster manapun maka akan ditandai sebagai *noise*.
 - f. Membuat plot klaster DBSCAN yang terbentuk.
 - g. Melakukan uji validitas menggunakan DBI, *C-Index*, dan *Dunn Index* kemudian dapat menjelaskan mengenai klaster yang terbentuk.

Davies-Bouldin Index merupakan metode evaluasi yang menghitung rata-rata nilai setiap titik pada himpunan data. Perhitungan nilai setiap titik adalah jumlah nilai *compactness* (kepadatan) yang dibagi dengan jarak antara kedua titik pusat klaster (*centroid*) sebagai *separation* (pemisah), menurut (Bates & Kalita, 2016). Jumlah klaster terbaik dapat dilihat dari hasil DBI yang semakin kecil (Khairati, A., G., & B., 2019). Langkah-langkah untuk menghitung DBI, meliputi:

- a. Menghitung nilai *Sum of Square Within cluster* (SSW), yaitu nilai yang digunakan untuk mengetahui kohesi dalam sebuah klaster ke-*i*. Rumus perhitungan nilai SSW yaitu:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)^2 \quad (4)$$

dimana m_i adalah jumlah data pada klaster ke-*i* dan $d(x_i, c_i)$ adalah jarak antara setiap objek ke-*i* terhadap pusat klaster ke-*i*.

- b. Menghitung nilai *Sum Between cluster* (SB), yaitu nilai yang digunakan untuk mengetahui separasi antar klaster. Rumus perhitungan nilai SB yaitu:

$$SB_{i,j} = d(c_i, c_j) \quad (5)$$

dimana $d(c_i, c_j)$ adalah jarak antara pusat klaster ke-*i* terhadap pusat klaster ke-*j*.

- c. Menghitung nilai rasio (R), yaitu nilai yang digunakan untuk mengetahui perbandingan antara klaster ke-*i* dengan klaster ke-*j*. Rumus perhitungan nilai rasio yaitu:

$$R_{i,j} = \frac{SSW_i + SS_j}{SSB_{i,j}} \quad (6)$$

- d. Setelah menghitung nilai rasio, kemudian nilai rasio tersebut digunakan untuk menghitung nilai *Davies-Bouldin Index* (DBI). Rumus perhitungan nilai DBI yaitu:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (7)$$

dimana k adalah jumlah kluster dan $R_{i,j}$ adalah rasio kluster ke- i dan kluster ke- j .

C-Index merupakan validasi kluster internal yang menyatakan ukuran jarak antar kluster dan jarak dalam kluster menurut (Nahdliyah, Tatik, & Alan, 2019). Perhitungan metode ini dengan cara membandingkan selisih jumlah antar objek di dalam tiap kluster dan jumlah minimum jarak antar objek, dengan selisih jumlah maksimum dan minimum jarak antar objek. Jumlah kluster terbaik dapat dilihat dari hasil *C-Index* yang semakin kecil. Adapun rumus perhitungan *C-Index* adalah sebagai berikut:

$$C_{Index} = \frac{S_W - S_{min}}{S_{max} - S_{min}} ; S_{min} \neq S_{max} \quad (8)$$

Keterangan:

S_W : Jumlah jarak antar objek di dalam tiap kluster

S_{min} : Jumlah jarak terkecil antara semua pasang objek pada seluruh kumpulan

S_{max} : Jumlah jarak terbesar antara semua pasang objek pada seluruh kumpulan

Dunn Index menghitung rasio antar jarak kluster minimum sebagai *separation* (pemisah) dan ukuran kluster maksimum sebagai *compactness* (kepadatan) menurut (Khairati, A., G., & B., 2019). Jumlah kluster terbaik dapat dilihat dari hasil *Dunn Index* yang semakin besar. Adapun rumus perhitungan *Dunn Index* adalah sebagai berikut:

$$Dunn\ Index = \frac{\min d(C_i, C_j)}{\max(diam(C_k))} \quad (9)$$

dimana $\min d(C_i, C_j)$ mengukur jarak minimum antar kluster dan $\max(diam(C_k))$ merupakan diameter kluster maksimum.

- 7) Melakukan interpretasi hasil kluster menggunakan metode terbaik.

C. Metode Analisis

1) K-Means

K-Means berusaha mempartisi data, dimana data dengan kesamaan karakteristik akan dikelompokkan ke dalam kluster yang sama, sedangkan kelompok dengan perbedaan karakteristik akan dikelompokkan ke dalam kluster yang lain (Suhartini & Ria, 2021). Objek-objek yang berada pada kluster yang sama memiliki tingkat homogenitas yang tinggi. Sementara itu, antar objek yang berada dalam suatu kluster dengan objek yang berada pada kluster lain memiliki tingkat heterogenitas tinggi.

2) Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Menurut Ashari **Invalid source specified.**, DBSCAN adalah algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data. DBSCAN menggunakan dua parameter yaitu *MinPts* dan *epsilon* (ϵ). Konsep kepadatan yang dimaksud dalam DBSCAN adalah minimal banyaknya data atau yang biasa disebut dengan *MinPts*, yang berada dalam radius atau *eps* (ϵ) dari setiap data. Konsep

kepadatan seperti ini menghasilkan tiga macam status dari setiap data, yaitu inti (*core*), batas (*border*), dan *noise*.

3. HASIL DAN PEMBAHASAN

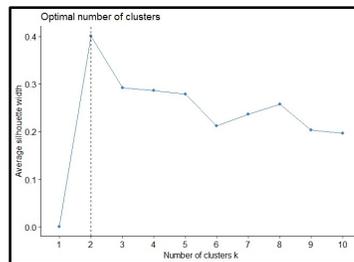
A. Pengelompokan Menggunakan Metode K-Means

Pengelompokan menggunakan metode *K-Means* dilakukan dengan dua pengukuran jarak, yaitu jarak *Euclidean* dan *Manhattan*.

1) K-Means dengan Jarak Euclidean

a. Menentukan Jumlah Kluster Optimal

Penentuan jumlah kluster optimal pada metode *K-Means* dengan jarak *Euclidean* menggunakan metode *silhouette width*, menunjukkan *average silhouette width* tertinggi terletak pada titik kedua, yang artinya jumlah kluster optimal sebanyak 2 kluster.



Gambar 1. Hasil *Silhouette Width* dengan Jarak *Euclidean*

b. Anggota Kluster yang Terbentuk

Penentuan jumlah kluster optimal sebanyak 2 kluster menghasilkan jumlah kabupaten dan kota di setiap klasternya seperti yang disajikan pada “Tabel 1”:

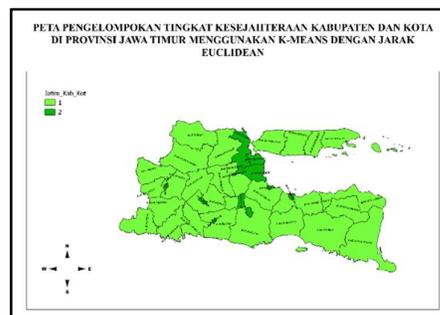
Tabel 1. Jumlah Anggota Kluster *K-Means* dengan Jarak *Euclidean*

Kluster	1	2
Jumlah	27	11

Hasil pada “Tabel 1” menunjukkan kluster terbentuk sebanyak 2 kluster dengan jumlah anggota pada kluster 1 sebanyak 27 kabupaten dan kluster 2 sebanyak 11 kabupaten dan kota.

c. Peta Kluster

Adapun peta hasil pengelompokan metode *K-Means* dengan jarak *Euclidean* disajikan pada “Gambar 2”:

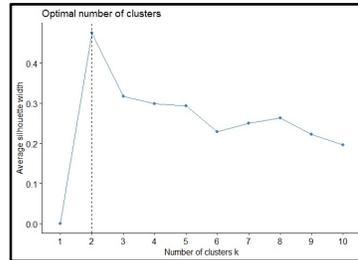


Gambar 2. Peta *K-Means* dengan Jarak *Euclidean*

2) *K-Means dengan Jarak Manhattan*

a. Menentukan Jumlah Kluster Optimal

Penentuan jumlah kluster optimal pada metode *K-Means* dengan jarak *Manhattan* menggunakan metode *silhouette width*, menunjukkan *average silhouette width* tertinggi terletak pada titik kedua, yang artinya jumlah kluster optimal sebanyak 2 kluster.



Gambar 3. Hasil *Silhouette Width* dengan Jarak *Manhattan*

b. Anggota Kluster yang Terbentuk

Penentuan jumlah kluster optimal sebanyak 2 kluster menghasilkan jumlah kabupaten dan kota di setiap klasternya seperti yang disajikan pada “Tabel 2”:

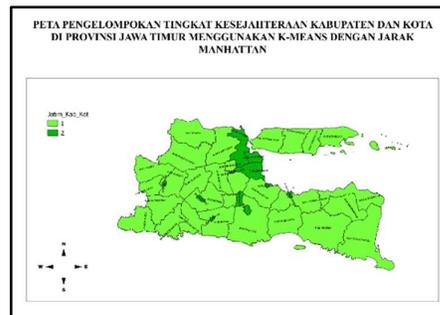
Tabel 2. Jumlah Anggota Kluster *K-Means* dengan Jarak *Manhattan*

Kluster	1	2
Jumlah	27	11

Hasil pada “Tabel 2” menunjukkan kluster terbentuk sebanyak 2 kluster dengan jumlah anggota pada kluster 1 sebanyak 27 kabupaten dan kluster 2 sebanyak 11 kabupaten dan kota.

c. Peta Kluster

Adapun peta hasil pengelompokan metode *K-Means* dengan jarak *Manhattan* disajikan pada “Gambar 4”:



Gambar 4. Peta *K-Means* dengan Jarak *Manhattan*

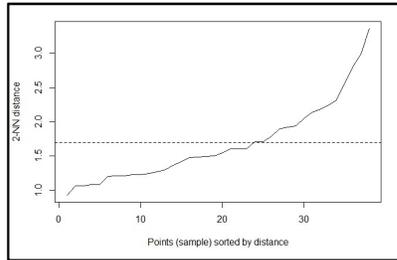
B. Pengelompokan Menggunakan Metode DBSCAN

Pengelompokan menggunakan metode DBSCAN dilakukan dengan dua pengukuran jarak, yaitu jarak *Euclidean* dan *Manhattan*.

1) *DBSCAN dengan Jarak Euclidean*

a. Inisialisasi Parameter

Penelitian ini menggunakan nilai *MinPts* sebesar 2, sedangkan nilai *eps* dengan jarak *Euclidean* yang ditentukan dari plot *K-Nearest Neighbor* (KNN) dan *Golden Section*.



Gambar 5. *K-Nearest Neighbor Plot* dengan Jarak *Euclidean*

Nilai *eps* yang dihasilkan dengan plot KNN sebesar 1,70. Kemudian dari hasil tersebut, ditarik nilai dari rentang 1 hingga 2 untuk dilakukan optimasi menggunakan *Golden Section* dengan toleransi sebesar 1%, karena pembentukan siku terjadi dari rentang tersebut. Iterasi *Golden Section* menghasilkan nilai optimal sebesar 2,00 maka *eps* yang digunakan untuk DBSCAN dengan jarak *Euclidean* ini sebesar 2.

b. Anggota Klaster yang Terbentuk

Jumlah klaster yang terbentuk menggunakan *MinPts* = 2 dan *eps* = 2, seperti yang disajikan pada “Tabel 3”:

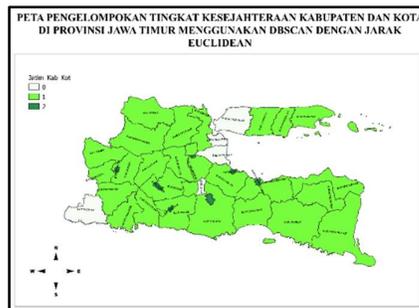
Tabel 3. Jumlah Anggota Klaster DBSCAN dengan Jarak *Euclidean*

Klaster	0	1	2
Jumlah	5	27	11

Hasil pada “Tabel 3” menunjukkan klaster terbentuk sebanyak 2 klaster dan 5 *noise* dengan jumlah anggota pada klaster 1 sebanyak 26 kabupaten, klaster 2 sebanyak 7 kota, dan *noise* sebanyak 5 kabupaten/kota.

c. Peta Klaster

Adapun peta hasil pengelompokan metode DBSCAN dengan jarak *Euclidean* disajikan pada “Gambar 6”:

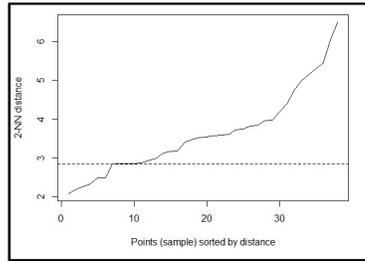


Gambar 6. Peta DBSCAN dengan Jarak *Euclidean*

2) *DBSCAN* dengan Jarak *Manhattan*

a. Inisialisasi Parameter

Penelitian ini menggunakan nilai *MinPts* sebesar 2, sedangkan nilai *eps* dengan jarak *Manhattan* yang ditentukan dari plot *K-Nearest Neighbor* (KNN) dan *Golden Section*.



Gambar 7. *K-Nearest Neighbor Plot* dengan Jarak *Manhattan*

Nilai *eps* yang dihasilkan dengan plot KNN sebesar 2,85. Kemudian dari hasil tersebut, ditarik nilai dari rentang 2,45 hingga 2,85 untuk dilakukan optimasi menggunakan *Golden Section* dengan toleransi sebesar 1%, karena pembentukan siku terjadi dari rentang tersebut. Iterasi *Golden Section* menghasilkan nilai optimal sebesar 2,85 maka *eps* yang digunakan untuk DBSCAN dengan jarak *Manhattan* ini sebesar 2,85.

b. Anggota Kluster yang Terbentuk

Jumlah kluster yang terbentuk menggunakan $MinPts = 2$ dan $eps = 2,85$ seperti yang disajikan pada “Tabel 4”:

Tabel 4. Jumlah Anggota Kluster DBSCAN dengan Jarak *Euclidean*

Kluster	0	1	2	3	4
Jumlah	19	12	2	3	2

Hasil pada “Tabel 4” menunjukkan kluster terbentuk sebanyak 4 kluster dan 19 *noise* dengan jumlah anggota pada kluster 1 sebanyak 12 kabupaten, kluster 2 sebanyak 2 kabupaten, kluster 3 sebanyak 3 kota, kluster 4 sebanyak 2 kota, dan *noise* sebanyak 19 kabupaten/kota.

c. Peta Kluster

Adapun peta hasil pengelompokan metode DBSCAN dengan jarak *Manhattan* disajikan pada “Gambar 8”:



Gambar 8. Peta DBSCAN dengan Jarak *Manhattan*

Pengelompokan DBSCAN dengan jarak *Manhattan* masih menghasilkan *noise* yang banyak, sehingga hasil pengelompokannya kurang bermanfaat. Oleh karena itu, dilakukan pengujian ulang dengan melakukan *trial* terhadap nilai parameter *epsilon*-nya, yaitu $eps = 4$. Percobaan tersebut menghasilkan 2 kluster dan 5 *noise*, dengan pengelompokan sama seperti DBSCAN dengan jarak *Euclidean*.

C. Perbandingan Metode K-Means dan DBSCAN

Setelah diperoleh hasil pengelompokan dari masing-masing metode, maka dilakukan perbandingan berdasarkan uji validitas, seperti berikut:

Tabel 5. Perbandingan Hasil Validitas Pengelompokan

No	Metode	Perhitungan Jarak	Parameter	Jumlah Klaster	Nilai Validitas		
					DBI	C-Index	Dunn Index
1	K-Means	Euclidean	$k = 2$	2 klaster	0,987	0,109	0,294
	K-Means	Manhattan	$k = 2$	2 klaster	0,367	0,073	0,269
2	DBSCAN	Euclidean	$MinPts = 2$	2 klaster + 5 noise	0,770	0,088	0,420
			$Eps = 2$				
	DBSCAN	Manhattan	$MinPts = 2$	4 klaster + 19 noise	0,434	0,025	0,491
			$Eps = 2,85$				
DBSCAN	Manhattan	$MinPts = 2$	2 klaster + 5 noise	0,284	0,073	0,382	
		$Eps = 4$					

Hasil validitas yang disajikan **Tabel 5** memperlihatkan bahwa berdasarkan masing-masing perhitungan jarak, metode pengelompokan *Density-Based Spatial Clustering of Applications with Noise* lebih baik dibandingkan dengan metode *K-Means*.

Nilai *C-Index* terendah dan *Dunn Index* tertinggi dihasilkan oleh metode DBSCAN menggunakan jarak *Manhattan* dengan $eps = 2,85$, dengan klaster yang terbentuk sebanyak 4 klaster dan 19 *noise*. Namun, metode terbaik tidak hanya dilihat dari nilai validitasnya saja. Oleh karena itu, metode DBSCAN menggunakan jarak *Manhattan* dengan $eps = 4$ lebih baik dibandingkan dengan $eps = 2,85$ karena menghasilkan lebih sedikit *noise* dengan DBI yang paling rendah daripada metode lain.

D. Cluster Profiling Metode Terbaik

Hasil terbaik pengelompokan kabupaten dan kota di Provinsi Jawa Timur menggunakan metode DBSCAN menggunakan jarak *Manhattan*, memiliki karakteristik dari masing-masing klasternya sebagai berikut:

Tabel 6. Cluster Profiling Metode Terbaik

No	Variabel	Klaster 0	Klaster 1	Klaster 2	Rata-rata
1	KP (Jiwa/km ²)	2904	815	5334	3017,67
2	AK (Jiwa)	753420	668878	158028	526775,33
3	TPAK (%)	71,0	71,2	68,3	70,17
4	TPT (%)	7,55	4,82	7,23	6,53
5	RLS (Tahun)	8,72	7,26	9,92	8,63
6	AHH (Tahun)	72,6	71,2	72,7	72,17
7	PP (Rupiah)	1286451	918897	1391347	1198898,33
8	PPM (%)	9,92	12,5	6,46	9,63

Hasil pada “**Tabel 6**” menunjukkan klaster 1 memiliki karakteristik tingkat kesejahteraan yang rendah, klaster 2 memiliki karakteristik tingkat kesejahteraan yang tinggi, dan klaster 0 atau *noise* memiliki karakteristik tingkat kesejahteraan yang sedang. Namun, daerah pada klaster 0 tidak dapat disebut sebagai kelompok karena memiliki karakteristik lain dan tidak memiliki kerapatan yang tinggi dengan daerah lain untuk membentuk suatu klaster.

4. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan mengenai perbandingan hasil pengelompokan kabupaten dan kota di Provinsi Jawa Timur berdasarkan tingkat kesejahteraan menggunakan metode *K-Means* dan *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*, maka dapat diambil beberapa kesimpulan antara lain:

- a. Pengelompokan tingkat kesejahteraan kabupaten dan kota di Provinsi Jawa Timur menggunakan metode *K-Means* dengan jarak *Euclidean* dan *Manhattan* menghasilkan 2 klaster berdasarkan hasil plot metode *silhouette width*.
- b. DBSCAN dengan jarak *Euclidean* menggunakan parameter $MinPts = 2$ dan $eps = 2$ menghasilkan 2 klaster dan 5 *noise*. DBSCAN dengan jarak *Manhattan* menggunakan parameter $MinPts = 2$ dan $eps = 2,85$ menghasilkan 4 klaster dan 19 *noise*. Sementara itu, DBSCAN dengan jarak *Manhattan* menggunakan parameter $MinPts = 2$ dan $eps = 4$ menghasilkan 2 klaster dan 5 *noise* dengan pengelompokan yang sama seperti DBSCAN dengan jarak *Euclidean*.
- c. Metode DBSCAN dengan jarak *Manhattan* menggunakan $MinPts = 2$ dan $eps = 4$ merupakan metode terbaik karena menghasilkan *noise* yang lebih sedikit dan DBI terkecil, dimana klaster 1 memiliki karakteristik tingkat kesejahteraan yang rendah, klaster 2 memiliki karakteristik tingkat kesejahteraan yang tinggi, dan klaster 0 atau *noise* merupakan daerah yang memiliki karakteristik tingkat kesejahteraan yang sedang.
- d. Dalam hal ini metode DBSCAN dengan jarak *Manhattan* lebih baik daripada *K-means* karena data yang digunakan mengandung *noise*, *high dimensionality*, dan juga hasil validitas DBI, *C-Index*, *Dunn Index* diperoleh hal yang serupa.

Berdasarkan hasil pengelompokan tersebut, diharapkan agar Pemerintah Provinsi Jawa Timur dapat memeriksa kembali pemerataan kesejahteraan di Provinsi Jawa Timur karena masih terdapat daerah-daerah yang mempunyai kesejahteraan rendah, seperti Kabupaten Sampang.

5. DAFTAR PUSTAKA

- Ana, E., Chamidah, N., Andriani, P., & Lestari, B. (2019). Pemodelan Faktor Risiko Hipertensi Menggunakan Linier Lokal Regresi Logistik Nonparametrik Aditif. *Jurnal Fisika*, 1-9.
- Anonim. (2022, Maret 6). Retrieved Oktober 12, 2021, from worldometers.info: <https://www.worldometers.info/world-population/#top20>
- Basuki, A. T. (2017). *Regresi Logistik*. Yogyakarta: Universitas Muhammadiyah Yogyakarta.
- Bates, A., & Kalita, J. (2016, March 4-5). Counting clusters in twitter posts. doi:<https://doi.org/10.1145/2905055.2905>
- BPS. (2021). Retrieved from Badan Pusat Statistik: <https://bps.go.id>
- Fithriasari, K., Soehardjoepri, & Iriawan, N. (2018). Generalized Additive Logistic Pada Pemodelan Faktor-faktor Yang Mempengaruhi Keuntungan PT. PDC. *Jurnal Inferensi*, 1(1), 45-47.
- IDM. (2021). Retrieved from IDM: idm.kemendesa.go.id
- Kartikawati, S., & Ambarwati, A. N. (2017). Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode Ward's. *Prosiding Seminar Nasional Pendidikan, Sains, Dan Teknologi*, 284-290.
- Khairati, A. F., A., A. A., G., F. H., & B., D. H. (2019). Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. *PRISMA*, 2, 161-170.
- Nahdliyah, M. A., Tatik, W., & Alan, P. (2019). Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index. *Jurnal Gaussian*, 8(2), 161-170.
- Newita, T. D., & Martha, S. (2018). Perbandingan Model Regresi Logistik dan Model Regresi Beta Binomial dalam Kasus Kredit Macet. *Jurnal Bimaster*, 7(4), 255-260.

- Nugraheny, D. (2015). Metode Nilai Jarak Guna Kesamaan atau Kemiripan Ciri Suatu Citra (Kasus Deteksi Awan Cumulonimbus Menggunakan Principal Component Analysis). *Jurnal Angkasa*, 7(2), 21-30.
- Risman, Syaripuddin, & Suyitno. (2019). Implementasi Metode DBSCAN pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat. *Prosiding SNMSA*, 22-28. Retrieved September 25, 2021, from <https://jurnal.fmipa.unmul.ac.id>
- Suhartini, & Ria, Y. (2021, Januari). Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur. *Jurnal Informatika dan Teknologi*, 4(1), 39-50. doi:<https://doi.org/10.29408/jit.v4i1.2986>
- Sukarsa, I. K., & Kencana, E. N. (2016). Generalized Additive Models (GAMs): Suatu Alternatif Pendeteksian Pengaruh Non-Linear dalam Analisis Regresi Berganda. *Jurnal Matematika*, 1-7