

# PENGELOMPOKAN KABUPATEN/KOTA DI PULAU KALIMANTAN PADA TAHUN 2020 DAN 2021 BERDASARKAN INDEKS PEMBANGUNAN MANUSIA MENGGUNAKAN METODE ALGORITMA ST-DBSCAN

## (*GROUPING OF DISTRICTS/CITIES ON KALIMANTAN IN 2020 AND 2021 BASED ON THE HUMAN DEVELOPMENT INDEX USING THE ST-DBSCAN ALGORITHM METHOD*)

Binda Aprilia Suryani\*, Memi Nor Hayati\*, Surya Prangga\*

\*Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman  
Jalan Penajam, Samarinda, Kalimantan Timur  
Email: bindaaprilia@gmail.com

Diterima: 3 Mei 2023; Direvisi: 9 Agustus 2023; Disetujui: 14 Agustus 2023

### ABSTRAK

*Clustering* merupakan suatu teknik menganalisis pengelompokan berbeda terhadap data. Rata-rata studi di *Clustering* fokus untuk menemukan pengelompokan dari data biasa (data non-spasial dan non-temporal), sehingga tidak praktis digunakan untuk mengelompokkan data spasial-temporal dirancang dalam banyak aplikasi seperti Sistem Informasi Geografis, Pencitraan Medis dan Perkiraan Cuaca. *Spatial Temporal-Density Based Spatial Clustering of Applications with Noise* (ST-DBSCAN) adalah algoritma pengelompokan berbasis kepadatan (*density*) yang memiliki kemampuan untuk mencari pengelompokan berdasarkan data spasial, data temporal, dan data non-spasial dari objek. Tujuan dari penelitian ini adalah untuk mengetahui nilai *Silhouette Coefficient* (*SC*) yang optimal dan hasil pengelompokan data IPM pada 56 Kabupaten/Kota di Pulau Kalimantan. Perhitungan nilai *SC* melibatkan parameter *Eps1* untuk data spasial, *Eps2* untuk data temporal dan *MinPts*. Nilai Parameter *Eps1* dan *MinPts* disimulasikan secara *trial and error* untuk menghasilkan nilai *SC* terbesar. Nilai parameter *Eps1* = 1 sampai dengan *Eps1* = 5 *Eps2* = 2 dan *MinPts* = 4 sampai dengan *MinPts* = 6. Berdasarkan hasil nilai *SC* yang tertinggi pada pengelompokan kabupaten/kota di Pulau Kalimantan menggunakan algoritma ST-DBSCAN adalah 0,324 yaitu terbentuk sebanyak 2 *cluster* dengan *cluster* pertama beranggotakan 42 kabupaten/kota dan *cluster* nol atau *outlier* beranggotakan 14 Kabupaten/Kota. Hasil pengelompokan tersebut dengan metode ST-DBSCAN ada 14 kabupaten/kota di Pulau Kalimantan Barat termasuk kelompok rendah dan dapat menjadi perhatian khusus pemerintah dalam meningkatkan IPM.

**Kata kunci:** ST-DBSCAN, IPM, *Silhouette Coefficient*, Pulau Kalimantan

### ABSTRACT

*Clustering* is a technique analyzing different groupings of data. Most studies in clustering focus on discovering clusters from ordinary data (non-spatial and non-temporal data), so they are impractical to use for clustering spatial-temporal data can be used in many applications such as geographic information systems, medical imaging, and weather forecasting. *Spatial Temporal-Density Based Spatial Clustering of Applications with Noise* (ST-DBSCAN) is a density-based clustering algorithm that has the ability to search for groupings based on spatial data, temporal data, and non-spatial data from the object. The purpose of this research was to know the value of the *Silhouette Coefficient* (*SC*) of optimal and the results of grouping HDI data in 56 districts/cities on in Kalimantan. The calculation of *SC* value involves *Eps1* parameter for spatial data, *Eps2* for temporal data and *MinPts*. The values of Parameter *Eps1* and *MinPts* were simulated by trial and error to produce the largest *SC* value. Parameter value *Eps1* = 1 to *Eps1* = 5 *Eps2* = 2 and *MinPts* = 4 to *MinPts* = 6. Based on the results of the highest *SC* value in the grouping of districts/cities in

*Kalimantan using the ST-DBSCAN algorithm was 0.324, which was formed as many as 2 clusters with the first cluster consisted of 42 regencies/cities and cluster zero or outlier consisted of 14 regencies/cities. The results of this grouping using the ST-DBSCAN method show that 14 districts/cities on West Kalimantan Island are in the low group and can be of particular concern to the government in increasing the HDI.*

**Keywords:** ST-DBSCAN, HDI, Silhouette Coefficient

## PENDAHULUAN

Data mining atau kadang disebut juga *Knowledge Discovery in Databases* (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar (Siregar & Puspabhuana, 2017). Salah satu teknik data mining yaitu melakukan pengelompokan (*Clustering*). *Clustering* adalah suatu teknik pengelompokan sejumlah data yang bersesuaian tanpa pengetahuan yang mendalam tentang kelompok tersebut. *Clustering* data dapat dibagi kedalam empat kelompok besar yaitu (1) metode partisi atau *Partitioning Methods*, (2) metode hirarki atau (*Hierarchical Methods*), (3) metode berbasis kerapatan atau *Density-Based Methods*, (4) metode berbasis grid (*Grid Based Methods*) (Irwansyah & Faisal, 2015).

*Density-Based Methods* merupakan *cluster* sebagai suatu area yang berisi objek-objek yang padat, yang dipisahkan oleh area yang memiliki kepadatan rendah atau merepresentasikan *noise*. Salah satu algoritma dari metode ini adalah ST-DBSCAN yang merupakan algoritma pengelompokan berbasis kepadatan (*density*) yang memiliki kemampuan untuk mencari pengelompokan berdasarkan data spasial, data temporal, dan data non-spasial dari objek. Kelebihan dari metode ST-DBSCAN adalah lebih fleksibel dengan ukuran data yang besar karena merupakan teknik yang berbasis kepadatan. (Birant & Kut, 2007).

Pulau Jawa memiliki pembangunan infrastruktur dan kualitas sumber daya manusia yang baik sedangkan Pulau Kalimantan masih membangun infrastruktur termasuk membangun IKN yang baru sehingga dapat mempengaruhi nilai IPM di Pulau Kalimantan. Pembangunan manusia di Kalimantan Tengah terus mengalami kemajuan, yang ditandai dengan terus meningkatnya nilai IPM. Pada tahun 2018, IPM Kalimantan Tengah telah mencapai 70,91 dibandingkan IPM tahun 2017 yang sebesar 69,79. Sementara di tingkat kabupaten dan kota bahwa pada 2017 terdapat 4 kabupaten dan kota yang termasuk kategori tinggi dengan IPM antara 70,23 hingga 80,34. Sedangkan untuk kategori sedang terdapat 9 daerah dengan IPM antara 67,04 hingga 69,73. Pada daerah terendah yakni Kabupaten Seruyan dengan IPM 67,04 sedangkan IPM yang tertinggi yakni di Kota Palangka Raya dengan IPM mencapai 80,34 dan IPM untuk daerah lainnya adalah Korawaringin Barat sebesar 72,46, Kotawaringin Timur sebesar 70,56, Kapuas sebesar 68,68, Barito Selatan sebesar 69,73, Barito Utara sebesar 69,72, Sukamara sebesar 67,52, Lamandau sebesar 69,70, Pulang Pisau sebesar 67,54, Gunung Mas sebesar 70,23, Barito Timur sebesar 70,82 dan Murung Raya sebesar 67,56 (Diskominfo, 2019).

Peneliti sebelumnya Eriansyah dan Syaffullah (2018) melakukan penelitian dengan algoritma ST-DBSCAN dan K-Means untuk mengelompokkan IPM di pulau Jawa tahun 2014-2016. Penelitian ini membandingkan nilai Silhouette Coefficient (SC) dari dua algoritma tersebut dan didapatkan hasil bahwa metode K-means lebih baik dalam pengelompokan IPM dari pulau Jawa pada tahun 2014-2016.

Pada permasalahan diatas, terjadi ketimpangan dari nilai IPM di Kalimantan Tengah yaitu berada di Kota Palangka Raya sebesar 80,34 jauh lebih tinggi dari nilai IPM di Kabupaten Seruyan sebesar 67,04. Hal tersebut dapat dilakukan pengelompokan nilai IPM yang tersebar di Pulau Kalimantan untuk lebih fokus dalam perencanaan pembangunan manusia yang masih rendah.

## METODE

Teknik pengumpulan data yang digunakan adalah teknik pengumpulan data sekunder yaitu data diperoleh dari website Badan Pusat Statistika <https://bps.go.id/>. Data yang digunakan dalam penelitian ini adalah data Indeks Pembangunan Manusia tahun 2020 sampai dengan 2021 di 56 Kabupaten/Kota Pulau Kalimantan. Teknik sampling pada penelitian ini menggunakan teknik *purposive sampling* yaitu teknik penentuan sampel dengan pertimbangan tertentu (sugiyono, 2014), dimana dalam penelitian ini mempertimbangkan ketersediaan data terbaru.

Penelitian ini menggunakan ST-DBSCAN dengan mengikuti jurnal dari Eriansyah dan Syaffullah (2018). Analisis data dilakukan dengan beberapa tahap dengan menggunakan aplikasi *software R studio* versi 1.10.153. Berikut adalah tahap-tahapannya:

1. Melakukan statistika deskriptif untuk data temporal yaitu data IPM tahun 2020 dan 2021.
2. Pengelompokan data dengan menerapkan metode ST-DBSCAN.
  - a. Menentukan parameter *Eps* dan *MinPts*, dimana *Eps* adalah radius atau jarak data pengamatan satu ke data pengamatan lainnya dan *MinPts* adalah minimal titik (data pengamatan) yang masuk ke dalam radius atau jarak *Eps*.
  - b. Kemudian menghitung jarak antar data pengamatan pada data temporal (data IPM tahun 2020 dan 2021) dan data spasial (data *longitude* dan data *latitude*) menggunakan persamaan berikut:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{im} - x_{jm})^2}$$

dengan  $i$  dan  $j$  adalah dua objek data yang memiliki  $m$  atribut bernilai numerik, yang dinyatakan sebagai  $i = (x_{i1}, x_{i2}, \dots, x_{im})$  dan  $j = (x_{j1}, x_{j2}, \dots, x_{jm})$  dengan:

- $d_{ij}$  : Jarak antara objek ke- $i$  dan ke- $j$   
 $x_{il}$  : Data dari objek ke- $i$  pada variabel ke- $l$   
 $x_{jl}$  : Data dari objek ke- $j$  pada variabel ke- $l$

- c. Menentukan titik-titik yang *density-reachable* (titik yang dapat dicapai) terhadap  $x_i$  yang berada dalam radius *Eps* yang dilihat berdasarkan hasil perhitungan jarak pada langkah b.
  - d. Lalu, mencari irisan titik-titik yang *density-reachable* dari data temporal dan data spasial.
  - e. Mengalokasikan titik-titik yang *density reachable* terhadap  $x_i$  menjadi satu cluster ( $C_p$ ).  
 Jika banyaknya titik yang *density-reachable* terhadap  $x_i$  lebih dari atau sama dengan *MinPts* maka  $x_i$  adalah *core point* sehingga terbentuk *cluster* dapat dilanjutkan ke titik lain sekitarnya yang merupakan titik *density reachable* terhadap  $x_i$ . Jika banyaknya titik yang *density reachable* terhadap  $x_i$  kurang dari *MinPts* adalah *border point*.
  - f. Menemukan titik-titik *density-reachable* terhadap  $x_i$  yang merupakan *core point* atau *border point* dengan mengulang langkah e. Proses berhenti bila tidak ada titik-titik baru yang *density reachable*.
  - g. Kemudian mengkategorikan data yang tidak masuk ke *cluster* manapun dikategorikan sebagai *outlier* atau *noise* dan masuk *cluster* 0.
3. Hasil cluster dari analisis ST-DBSCAN menghitung nilai *silhouette coefficient* untuk mengetahui hasil *cluster* terbaik.
  - a. Menghitung rata-rata jarak dari suatu data ke- $i$  dengan semua data yang berada pada satu *cluster* yang sama dengan persamaan:

$$a_i = \frac{1}{n_p - 1} \sum_{j=1}^{n_p-1} d_{i,j}, j \neq i$$

Dengan  $p = 1, 2, \dots, k$ .

- b. Menghitung rata-rata jarak dari suatu data ke- $i$  dengan semua data yang berada pada satu *cluster* yang berbeda dengan persamaan:

$$d_i(p) = \frac{1}{n_p} \sum_{j=1}^{n_p} d_{i,j}$$

c. Menghitung nilai  $SC_1(i)$  untuk setiap data ke- $i$  dengan persamaan:

$$SC_1(i) = \frac{b_i - a_i}{\max\{a_i, b_i\}}, i = 1, 2, \dots, n$$

d. Menghitung rata-rata nilai  $SC_2(p)$  dengan persamaan:

$$SC_2(p) = \frac{1}{n_p} \sum_{x_i \in C_p}^{n_p} SC_1(i)$$

e. Menghitung nilai  $SC$  global dengan persamaan:

$$SC = \frac{\sum_{p=1}^k n_p SC_2(p)}{\sum_{p=1}^k n_p}$$

f. Kemudian mendapatkan nilai *cluster* dari  $SC$  yang Optimal.

## HASIL DAN PEMBAHASAN

### Analisis Statistik Deskriptif

Analisis statistika deskriptif pada data IPM di 56 Kabupaten/Kota Pulau Kalimantan pada Tahun 2020 dan 2021 menggunakan nilai minimum, nilai maksimum, rata-rata dan simpangan baku. Hasil analisis deskriptif dapat dilihat pada Tabel 1.

Tabel 1. Statistika Deskriptif Data Penelitian

	IPM 2020	IPM 2021
Minimum	62,68	62,90
Maksimum	80,77	80,82
Rata-rata	70,379	70,764
Simpangan Baku	4,363	4,418

Berdasarkan Tabel 1 di atas menunjukkan bahwa IPM pada tahun 2020 di Pulau Kalimantan dapat diketahui nilai simpangan bakunya atau rata-rata jarak nilai IPM tersebar di Pulau Kalimantan sebesar 4,363. IPM tahun 2020 tertinggi terdapat di Provinsi Kalimantan Tengah berada di Kota Palangka Raya sebesar 80,77 termasuk kategori sangat tinggi sedangkan IPM tahun 2020 terendah terdapat di Provinsi Kalimantan Barat berada di Kota Kayong Utara sebesar 62,68 termasuk kategori sedang. Adapun diketahui rata-rata IPM tahun 2020 sebesar 70,379 yang termasuk kategori tinggi. IPM pada tahun 2021 di Pulau Kalimantan dapat diketahui nilai simpangan bakunya atau rata-rata jarak nilai IPM tersebar di Pulau Kalimantan sebesar 4,418 IPM tahun 2021 tertinggi terdapat di Provinsi Kalimantan Tengah berada di Kota Palangka Raya sebesar 80,82 termasuk kategori sangat tinggi sedangkan IPM tahun 2021 terendah terdapat di Provinsi Kalimantan Barat berada di Kota Kayong Utara sebesar 62,90 termasuk kategori sedang. Adapun diketahui rata-rata IPM tahun 2021 sebesar 70,764 yang termasuk kategori tinggi.

### Analisis Cluster dengan Algoritma ST-DBSCAN

#### 1. Penentuan Parameter

Pada penelitian ini nilai yang digunakan untuk parameter  $Eps1$  adalah 1, 2, 3, 4 dan 5. Kemudian nilai parameter  $Eps2$  yang digunakan adalah 2, karena waktu yang digunakan selama dua tahun yaitu dari tahun 2020 sampai tahun 2021. Nilai  $MinPts$  yang digunakan adalah 4, 5 dan 6. Hasil *cluster* yang terbentuk berdasarkan nilai-nilai parameter yang ditentukan nantinya akan dievaluasi menggunakan nilai *silhouette coefficient*.

## 2. Jarak Antar Data Pengamatan

Salah satu contoh perhitungan jarak Euclid data *longitude* dan data *latitude* adalah sebagai berikut:

$$\begin{aligned} d_{1,2_{SP}} &= \sqrt{(x_{1,1} - x_{2,1})^2 + (x_{1,2} - x_{2,2})^2} \\ &= \sqrt{(-1,159 - 1,930)^2 + (116,868 - 117,295)^2} \\ &= 3,118 \end{aligned}$$

∩

$$\begin{aligned} d_{55,56_{SP}} &= \sqrt{(x_{55,1} - x_{56,1})^2 + (x_{55,2} - x_{56,2})^2} \\ &= \sqrt{(-0,140 - (-2,206))^2 + (114,415 - 113,912)^2} \\ &= 2,127 \end{aligned}$$

Selanjutnya untuk perhitungan jarak Euclid data temporal (data IPM pada tahun 2020 dan 2021) adalah sebagai berikut:

$$\begin{aligned} d_{1,2_{TP}} &= \sqrt{(x_{1,1} - x_{2,1})^2 + (x_{1,2} - x_{2,2})^2} \\ &= \sqrt{(80,01 - 74,71)^2 + (80,71 - 74,72)^2} \\ &= 7,645 \end{aligned}$$

∩

$$\begin{aligned} d_{55,56_{TP}} &= \sqrt{(x_{55,1} - x_{56,1})^2 + (x_{55,2} - x_{56,2})^2} \\ &= \sqrt{(67,89 - 80,77)^2 + (67,89 - 80,82)^2} \\ &= 18,088 \end{aligned}$$

## 3. Penentuan *Density Reachable*

Langkah awal yang perlu dilakukan penentuan *density reachbale* adalah dengan melihat nilai jarak Euclid apakah sesuai terhadap *Eps1*, *Eps2* dan *MinPts* yang ditentukan. Sebagai contoh menggunakan parameter *Eps1* = 1, *Eps2* = 2 dan *MinPts* = 4 untuk data spasial pada titik  $x_1$  (Kota Balikpapan) terhadap titik  $x_2$  (Kabupaten Berau) bahwa hasil  $d_{1,2_{SP}}$  adalah 3,118 dapat diartikan titik  $x_1$  terhadap titik  $x_2$  bukan *density reachable* karena hasil jarak Euclid titik tersebut tidak berada dalam *Eps1* ( $d_{1,2_{SP}} > Eps1$ ) maka dapat disebut sebagai *outlier*. Untuk titik  $x_1$  (Kota Balikpapan) terhadap titik  $x_5$  (Kabupaten Kutai Kartanegara) bahwa hasil  $d_{1,5_{SP}}$  adalah 0,083 dapat diartikan titik  $x_1$  terhadap titik  $x_5$  adalah *density reachable* karena hasil jarak Euclid titik tersebut berada dalam *Eps1* ( $d_{1,5_{SP}} \leq Eps1$ ) dan dilanjutkan ke titik lainnya.

Berikutnya data temporal pada titik  $x_1$  (Kota Balikpapan) terhadap titik  $x_2$  (Kabupaten Berau) bahwa hasil  $d_{1,2_{TP}}$  adalah 7,645 dapat diartikan titik  $x_1$  terhadap titik  $x_2$  bukan *density reachable* karena hasil jarak Euclid titik tersebut tidak berada dalam *Eps2* ( $d_{1,2_{TP}} > Eps2$ ) maka dapat disebut sebagai *outlier*. Untuk titik  $x_1$  Kota Balikpapan) terhadap titik  $x_3$  (Kabupaten Bontang) bahwa hasil  $d_{1,3_{TP}}$  adalah 0,120 dapat diartikan titik  $x_1$  terhadap titik  $x_3$  adalah *density reachable* karena hasil jarak Euclid titik tersebut berada dalam *Eps2* ( $d_{1,3_{TP}} \leq Eps2$ ) dan dilanjutkan ke titik lainnya.

## 4. Penentuan Irisan Data Temporal dan Data Spasial

Setelah menemukan beberapa titik  $x_1$  yang *density reachable* dari data spasial dan data temporal, kemudian akan ditentukan irisannya atau mencari titik *density reachable* yang

sama dari data spasial dan data temporal ( $A \cap B = \{x_i | x_i \in A \text{ dan } x_i \in B\}$ ). Adapun contoh titik *density reachable* yang bukan irisan sebagai berikut:

Data spasial :  $d_{1,3_{SP}} = 3,118$  (Bukan *Density Reachable*)  
 Data temporal :  $d_{1,3_{TP}} = 0,120$  (*Density Reachable*)

Berdasarkan contoh diatas dapat dilihat data spasial dan data temporal tidak ada irisan atau kesamaan data. Data yang tidak memiliki irisan/kesamaan *density reachable* secara spasial dan temporal akan diabaikan atau tidak dihitung dalam jumlah titik minimum (*MinPts*). Sedangkan contoh titik *density reachable* yang irisan atau sama sebagai berikut:

Data spasial :  $d_{1,10_{SP}} = 0,718$  (*Density Reachable*)  
 Data temporal :  $d_{1,10_{TP}} = 0,112$  (*Density Reachable*)

Pada contoh diatas dapat dijelaskan bahwa data spasial dan data temporal memiliki irisan atau kesamaan data maka termasuk dalam hitungan jumlah titik minimum (*MinPts*).

#### 5. Penempatan Titik-Titik *Density Reachable* dalam Cluster

Diketahui bahwa jumlah titik *density reachable* pada data ke-1 (Kota Balikpapan) adalah sebanyak 2 titik *density reachable*, yakni Kota Balikpapan sendiri dan Kota Samarinda ( $x_{10}$ ) ( $\|N_{Eps}(x_1)\| < MinPts = 4$ ) maka titik  $x_1$  (Kota Balikpapan) dapat dianggap sebagai *outlier*.

Data ke-11 (Kabupaten Sambas) memiliki jumlah titik *density reachable* sebanyak 4 titik yakni Kabupaten Sambas sendiri, Kabupaten Bengkayang ( $x_{12}$ ), Kabupaten Landak  $x_{13}$  dan Kabupaten Mempawah ( $x_{14}$ ) ( $\|N_{Eps}(x_{11})\| \geq MinPts = 4$ ) maka  $x_{11}$  (Kabupaten Sambas) dapat dianggap sebagai titik *core* pertama yang berarti kita telah mendapatkan titik pusat untuk *cluster* pertama. Selanjutnya akan dicari titik-titik lainnya yang akan menjadi anggota *cluster* pertama. Adapun titik-titik lainnya yang berkaitan dengan kabupaten/kota yang menjadi titik *core* pertama seperti yang ditampilkan pada Tabel 2.

Tabel 2 Data Kabupaten/ Kota yang termasuk Cluster Pertama

Kabupaten/ Kota	Titik <i>Density Reachable</i>	Jumlah Minimum $\geq MinPts=4$	Core/Border
11	$x_{11}, x_{12}, x_{13}, x_{14}$	$4 \geq MinPts$	Core
12	$x_{12}, x_{11}$	$2 \geq MinPts$	Border
13	$x_{13}, x_{11}, x_{14}, x_{15}$	$4 \geq MinPts$	Core
14	$x_{14}, x_{11}, x_{13}$	$3 \geq MinPts$	Border
15	$x_{15}, x_{13}, x_{19}$	$3 < MinPts$	Border

#### Validasi Cluster

Pada bagian ini digunakan untuk melihat kualitas dan kekuatan dengan menggunakan metode *Silhouette Coefficient (SC)*. Untuk mengetahui kualitas dari hasil *cluster*. Setelah mendapatkan nilai *SC* untuk data spasial dilanjutkan pencarian nilai *SC* untuk data temporal. Nilai *SC* global ini yang akan dijadikan acuan untuk memilih kombinasi parameter optimal. Semakin mendekati 1, maka hasil pengelompokan yang dihasilkan dari kombinasi parameter parameter tersebut semakin baik. Nilai parameter untuk  $Eps1 = 1, 2, 3, 4$  dan  $5$ ,  $Eps2 = 2$  dan  $MinPts = 4, 5$  dan  $6$  untuk data spasial dan data temporal menghasilkan banyak *cluster* serta nilai *SC*.

Pada Tabel 3 dapat dilihat bahwa analisis ST-DBSCAN menghasilkan jumlah *cluster* yang bervariasi sebanyak 2 macam jumlah *cluster* yaitu 3 dan 2 *cluster*. Pada tabel diatas yang menghasilkan 3 *cluster* sebanyak 7 dengan hasil rata-rata *SC* berturut-turut yaitu -0,046; 0,181; 0,183; 0,162; 0,173; 0,196 dan 0,209, maka hasil rata-rata *SC* terbaik yang didapat dari 3 *cluster* yaitu 0,209 dengan parameter  $Eps1 = 5$ ,  $Eps2 = 2$  dan  $MinPts = 5$ . Yang menghasilkan 2 *cluster*

Pengelompokan Kabupaten/Kota di Pulau Kalimantan Pada Tahun 2020 dan 2021 Berdasarkan Indeks Pembangunan Manusia Menggunakan Metode Algoritma ST-DBSCAN  
**Binda Aprilia Suryani, Memi Nor Hayati, Surya Prangga**

sebanyak 8 dengan hasil rata-rata *SC* berturut-turut, yaitu -0,029; 0,079; 0,209; 0,261; 0,270; 0,270 0,323 dan 0,344, maka hasil rata-rata *SC* terbaik yang didapat dari 2 *cluster* yaitu 0,344 dengan parameter *Eps1* = 5, *Eps2* = 2 dan *MinPts*= 6. Berdasarkan hasil perhitungan rata-rata nilai *SC* dari setiap 2 macam jumlah *cluster* diambil yang tertinggi ialah 0,344 dari parameter *Eps1* = 5, *Eps2* = 2 dan *MinPts* = 6 dengan menghasilkan 2 *cluster* yaitu *cluster* 1 dan *cluster* 0.

Tabel 3 Hasil Nilai *SC* berserta Jumlah *Cluster* yang dihasilkan Parameter *Eps1*= 1, 2, 3, 4 dan 5, *Eps2*= 2  
*MinPts* = 4,5 dan 6

<i>Eps1</i>	<i>Eps2</i>	<i>MinPts</i>	Jumlah <i>Cluster</i>	Jumlah <i>Noise</i>	<i>SC Data Spasial</i>	<i>SC Data Temporal</i>	Rata-rata <i>SC</i>
1	2	4	3	37	0,024	-0,117	-0,046
1	2	5	2	44	0,056	-0,115	-0,029
1	2	6	2	47	-0,181	0,024	-0,079
2	2	4	2	21	0,297	0,121	0,209
2	2	5	3	21	0,167	0,194	0,181
2	2	6	3	22	0,179	0,187	0,183
3	2	4	2	14	0,440	0,081	0,261
3	2	5	2	16	0,417	0,124	0,270
3	2	6	2	16	0,417	0,124	0,270
4	2	4	3	6	0,449	-0,126	0,162
4	2	5	3	7	0,451	-0,106	0,173
4	2	6	2	13	0,503	0,144	0,323
5	2	4	3	4	0,553	-0,160	0,196
5	2	5	3	5	0,543	-0,126	0,209
5	2	6	2	10	0,636	0,051	<b>0,344</b>

Tabel 4. Hasil Pengelompokan Metode ST-DBSCAN dengan *SC* yang Terbaik

<b>Cluster</b>	<b>Titi-titik</b>	<b>Kabupaten/Kota</b>
0 (Outlier)	$x_1, x_2, x_3, x_{10}, x_{21}, x_{23}, x_{36}, x_{37}, x_{42}, x_{56}$	Kota Balikpapan, Kabupaten Berau, Kota Bontang, Kabupaten Kutai Timur, Kota Samarinda, Kabupaten Kapuas Hulu, Kota Pontianak, Kota Banjar Baru, Kabupaten Malinau, Kabupaten Bulungan, Kabupaten Tana Tidung, Kabupaten Nunukan, Kota Tarakan, Kota Palangka Raya
1	$x_4, x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{20}, x_{22}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{43}, x_{44}, x_{45}, x_{46}, x_{47}, x_{48}, x_{49}, x_{50}, x_{51}, x_{52}, x_{53}, x_{54}, x_{55}$	Kabupaten Kutai Barat, Kabupaten Kutai Kartanegara, Kabupaten Mahakam Ulu, Kabupaten Paser, Kabupaten Penajam Paser Utara, Kabupaten Sambas, Kabupaten Bengkayang, Kabupaten Landak, Kabupaten Mempawah, Kabupaten Sanggau, Kabupaten Ketapang, Kabupaten Sintang, Kabupaten Sekadau, Kabupaten Melawi, Kabupaten Kayong Utara, Kabupaten Kubu Raya, Kota Singkawang, Kabupaten Tanah Laut, Kabupaten Kota Baru, Kabupaten Banjar, Kabupaten Barito Juala, Kabupaten Tapin, Kabupaten Hulu Sungai Selatan, Kabupaten Hulu Sungai Tengah, Kabupaten Hulu Sungai Utara, Kabupaten Tabalong, Kabupaten Tanah bumbu, Kabupaten Balangan, Kota Banjarmasin, Kabupaten Kotawaringin Barat, Kabupaten Kotawaringin Timur, Kabupaten Kapuas, Kabupaten Barito Selatan, Kabupaten Barito Utara, Kabupaten Sukamara, Kabupaten Lamandau, Kabupaten Seruyan, Kabupaten Katingan, Kabupaten Pulang Pisau, Kabupaten Gunung Mas, Kabupaten Barito Timur, Kabupaten Murung Raya, Kota Palangka Raya

### Hasil Cluster Terbaik

Analisis *cluster* data IPM di Pulau Kalimantan dengan metode ST-DBSCAN mendapatkan *SC* yang optimal sebesar 0,344 dengan kombinasi parameter yaitu  $Eps1 = 5$   $Eps2 = 2$   $MinPts = 6$ . Dengan bantuan *software* R studio maka hasil pengelompokan dengan menggunakan metode ST-DBSCAN yang menghasilkan beranggotakan 46 kabupaten/kota yang termasuk dalam *cluster* 1 sedangkan 10 kabupaten/kota lainnya terdeteksi sebagai kabupaten/kota yang *outlier* dan lebih jelasnya dapat dilihat pada Tabel 4.

Pada Tabel 4. Dapat dilihat bahwa *cluster* 0 menunjukkan bahwa lebih banyak kabupaten/kota yang berasal dari Provinsi Kalimantan Timur dan Provinsi Kalimantan Utara dimana provinsi tersebut memiliki kabupaten/kota nilai IPM yang tinggi. Sedangkan, *cluster* 1 menunjukkan bahwa lebih banyak kabupaten/kota dari Provinsi Kalimantan Barat dan memiliki kabupaten/kota nilai IPM yang rendah. IPM di Pulau Kalimantan memiliki rata-rata yang rendah sehingga perlu ada pergerakan dari pemerintah untuk meningkat nilai IPM dalam perencanaan pembangunan dengan bijak dan meningkatkan Angka Harapan Hidup, Harapan Lama Sekolah dan Rata-rata Lama Sekolah.

### KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, maka kesimpulan yang dapat diambil adalah sebagai berikut

1. Nilai *Silhouette Coefficient* untuk validasi data hasil *clustering* IPM di Pulau Kalimantan dengan metode ST-DBSCAN adalah sebesar 0,341 yang menyatakan bahwa struktur *cluster* yang dihasilkan pada pengelompokan ini adalah *weak structure* (struktur *cluster* yang lemah/buruk), artinya sedikit sekali objek  $x_i$  yang berada dalam *cluster* yang tepat dan sisanya bergabung pada *cluster* yang lainnya.
2. Hasil *Cluster* yang diperoleh berdasarkan hasil nilai *SC* optimal pada pengelompokan Kabupaten/Kota berdasarkan IPM di Pulau Kalimantan dengan menggunakan metode ST-DBSCAN adalah sebanyak 2 *cluster* yaitu *cluster* 0 dan *cluster* 1. *Cluster* 0 beranggotakan 14 Kabupaten/kota sisanya masuk ke dalam *cluster* 1 yang beranggotakan 42 kabupaten/kota.

### REKOMENDASI

Pemerintah kabupaten/kota di pulau Kalimantan seharusnya harus lebih memperhatikan Indeks Pembangunan Manusia (IPM) dengan memberikan alokasi dan realisasi anggaran dibidang kesehatan dan pendidikan secara merata. Karena adanya pengaruh kenaikan dan penurunan nilai Indeks Pembangunan Manusia.

### DAFTAR PUSTAKA

- Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60(1), 208–221. <https://doi.org/10.1016/j.datak.2006.01.013>
- Badan Pusat Statistik Provinsi Kalimantan Timur. (2018). *Berita Resmi Statistik: Indeks Pembangunan Manusia Kalimantan Timur Tahun 2017*. Samarinda: Badan Pusat Statistik Provinsi Kalimantan Timur.
- Dinas Komunikasi dan Informatika Provinsi Kalimantan Utara. (2019). *Profil IPM Kalimantan Utara Tahun 2019*. Dinas Komunikasi dan Informatika Provinsi Kalimantan Utara.
- Irwansyah, E., & Faisal, M. (2015). *Advanced Clustering: Teori dan Aplikasi* (1st ed.). Deepublish.



- Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (1998). Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications. *Journal Kluwer Academic Publisher*, 2, 169–194. <https://doi.org/https://doi.org/10.1023/A:1009745219419>
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis* (1st ed.). Graha Ilmu.
- Siregar, A. M., & Puspabhuana, A. (2017). *Data Mining: Pengolahan Data Menjadi dengan Informasi RapidMiner* (A. K. Putra (ed.); 1st ed.). Kekata Publisher.
- Sugiyono. (2014). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Alfabeta.
- Suryabrata, S. (2003). *Metode Penelitian* (1st ed.). Rajawali Pers.