

Comparison of DBSCAN and *K-Means* Clustering for Grouping the Village Status in Central Java 2020

Komparasi DBSCAN dan *K-Means Clustering* pada Pengelompokan Status Desa di Jawa Tengah Tahun 2020

Mustika Putri M^{1*}, Cesaria Dewi^{2*}, Emban Permata Siam^{3*}
Gona Asri Wijayanti^{4*}, Nurfitri Aulia^{5*}, Rani Nooraeni^{6*}

Abstract

Covid-19 as a pandemic disaster has begun to shake the world economic order, including Indonesia. Indonesia's economic growth has contract since quarter II. Central Java Province is in the third place with the highest number of positive cases in Indonesia. The government still improving quality control over the implementation of village funds by observing classification of village status. The Ministry of Village has classified village status based on IDM value by calculating the average of 3 indices, IKS, IKL, and IKE. This research was conducted with the aim of classifying the village status using 2 clustering methods, DBSCAN and *K-Means*. Another purpose of this research is to find out whether the same village status also groups in the same cluster and to compare which one of the two grouping methods is better in classifying village status in Central Java Province in 2020 based on the Silhouette coefficient value.

Keywords: Village Status, DBSCAN clustering, *K-Means* clustering, Covid-19

Abstrak

Sejak Covid-19 dinyatakan sebagai bencana pandemi, tatanan ekonomi dunia mulai terguncang, tidak terkecuali Indonesia. Pertumbuhan ekonomi Indonesia berkontraksi sejak triwulan II. Provinsi Jawa Tengah menempati urutan ketiga dengan jumlah kasus positif tertinggi di Indonesia. Pemerintah terus berupaya meningkatkan pengendalian mutu atas pelaksanaan dana desa salah satunya dengan memperhatikan klasifikasi status desa. Kementerian Desa telah membuat klasifikasi status desa berdasarkan nilai IDM dengan menghitung rata-rata dari 3 Indeks yaitu IKS, IKL, dan IKE. Adapun penelitian ini dilakukan dengan tujuan mengklasifikasikan status desa dengan menggunakan 2 (dua) metode *clustering* yaitu DBSCAN dan *K-Means*. Tujuan lain dari penelitian ini adalah untuk mengetahui apakah status desa yang sama juga mengelompok pada *cluster* yang sama serta membandingkan dari kedua metode *clustering* tersebut manakah yang lebih baik dalam mengklasifikasikan status desa di Provinsi Jawa Tengah tahun 2020 berdasarkan nilai koefisien *Silhouette*.

Kata kunci: Status desa, DBSCAN *clustering*, *K-Means clustering*, Covid-19

*Politeknik Statistika STIS,

Email: cesariadewi1@gmail.com, epermatas@yahoo.com, gonawijayanti@gmail.com,
mustikaputri.m1@gmail.com, nurfitriaulia91@gmail.com, raninoor@stis.ac.id



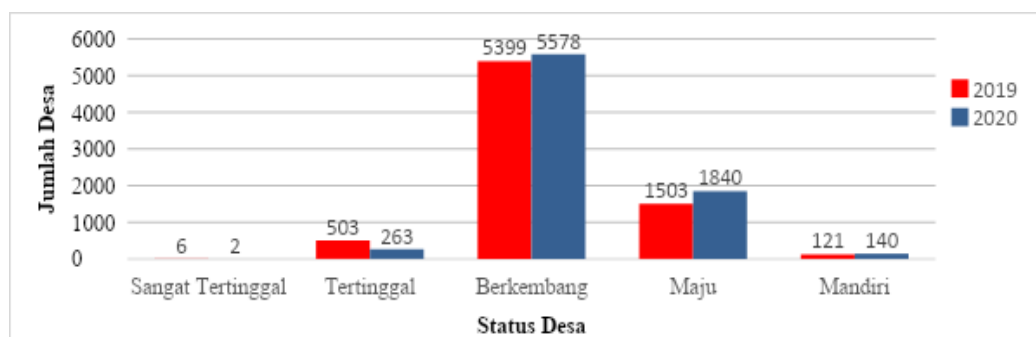
This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

1. PENDAHULUAN

Pada tahun 2020, hampir seluruh negara di dunia menghadapi pandemik Covid-19 termasuk di Indonesia. Pandemik tersebut pertama kali terkonfirmasi di Wuhan, China pada akhir Desember 2019 dan di Indonesia pada akhir Maret 2020. Pada tanggal 13 September 2020, Satgas Covid-19 melaporkan terdapat sekitar 214.746 orang terkonfirmasi Covid-19 [9]. Pandemik yang berlangsung telah menyebar ke berbagai aspek hingga mengganggu stabilitas di bidang kesehatan, perekonomian, dan sosial. Dalam bidang ekonomi, pemerintah mengeluarkan berbagai kebijakan untuk mempercepat pemulihan ekonomi nasional salah satunya dengan meningkatkan *quality control* pelaksanaan dana desa dalam pembangunan aksesibilitas, dan konektivitas sebagai upaya peningkatan pertumbuhan ekonomi. Untuk mendukung hal tersebut, pemerintah mengeluarkan UU No. 6 tahun 2014 dan PP No. 22 Tahun 2015 yang mengatur tentang pengalokasian dana desa dengan mempertimbangkan status kemajuan dan kemandirian desa.

Dalam upaya mencapai sasaran pembangunan untuk meningkatkan pertumbuhan ekonomi, dikembangkan Indeks Desa Membangun (IDM) yang tertuang dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 yaitu terwujudnya desa tertinggal menjadi desa berkembang sebanyak 10.000 desa dan terwujudnya desa berkembang menjadi desa mandiri sebanyak 5.000 desa [1]. Nilai IDM merupakan nilai yang dihitung berdasarkan nilai rata-rata dari 3 (tiga) indeks yakni IKS atau Indeks Ketahanan Sosial (pendidikan, kesehatan, modal sosial dan permukiman), IKE atau Indeks Ketahanan Ekonomi (keragaman produksi masyarakat, akses pusat perdagangan dan pasar, akses logistik, akses perbankan dan kredit, serta keterbukaan wilayah), dan IKL atau Indeks Ketahanan Lingkungan (kualitas lingkungan, bencana alam dan tanggap bencana). Nilai IDM tersebut selanjutnya dijadikan dasar dalam penentuan klasifikasi status desa, yakni desa sangat tertinggal, desa tertinggal, desa berkembang, desa maju, dan desa mandiri [5].

Provinsi Jawa Tengah dipilih dalam penelitian ini karena provinsi tersebut merupakan provinsi yang memiliki perkembangan jumlah desa yang besar setiap tahunnya. Selain itu Jawa Tengah juga menjadi provinsi dengan jumlah kasus Covid-19 terbanyak ketiga yang mencapai 17.460 kasus [8] dan mengakibatkan terjadi kontraksi yang cukup besar pada pertumbuhan ekonomi kuartal kedua sebesar -5.94 persen [2].



Sumber: PDDI Kementerian Desa, diolah

Gambar 1.1. Jumlah Desa Berdasarkan Status IDM Jawa Tengah Tahun 2019-2020

Berdasarkan grafik di atas, status desa di Provinsi Jawa Tengah mengalami peningkatan dari tahun 2019. Peningkatan itu diiringi pula oleh peningkatan jumlah desa yang disebabkan adanya pemekaran pada tahun 2020. Peningkatan jumlah desa dari tahun 2019 sebesar 7532 menjadi 7823 pada tahun 2020. Jumlah desa dengan status desa mandiri meningkat dari 121 menjadi 140,

jumlah desa dengan status maju juga mengalami peningkatan dari 1503 menjadi 1840, jumlah desa berkembang juga meningkat dari 5339 menjadi 5578, desa tertinggal menurun dari 503 menjadi 263 desa, dan jumlah desa dengan status sangat tertinggal turun dari 6 menjadi 2 desa.

Berdasarkan uraian diatas, menurut peneliti klasifikasi status desa yang telah dilakukan oleh Kementerian Desa relatif sederhana karena hanya ditentukan berdasarkan rata-rata IKS, IKE, dan IKL saja. Oleh karena itu, perlu dilakukan penelitian dengan metode lain untuk melihat apakah terdapat perbedaan hasil pengelompokan status desa yang telah dilakukan oleh Kementerian Desa. *Clustering* merupakan salah satu metode data mining jenis *unsupervised learning* (bersifat tanpa arahan) yang dapat digunakan untuk mencari dan mengelompokkan suatu objek berdasarkan kemiripan karakteristik suatu data dengan data lainnya. Beberapa metode *clustering* yang sering digunakan yaitu *K-Means*, *Hierarchical/Division clustering*, *Fuzzy clustering*, DBSCAN, dan lainnya. Penelitian ini memanfaatkan metode *clustering* yaitu *K-Means* dan DBSCAN dalam mengelompokkan desa-desa yang memiliki kemiripan karakteristik satu sama lain. Metode ini dilakukan untuk melihat hasil perbandingan berdasarkan nilai koefisien *silhouette* antara penentuan status desa yang diberikan oleh kementerian desa dengan status desa hasil analisis. Metode tersebut dapat digunakan untuk memperoleh kelompok-kelompok desa. Setiap desa yang berada dalam satu kelompok berarti memiliki karakteristik yang mirip sedangkan antar kelompok memiliki karakteristik yang berbeda. Selain itu, metode ini juga lebih tepat digunakan untuk tipe data numerik [5].

Untuk Metode DBSCAN, proses pengelompokan ditentukan berdasarkan nilai *Minimal Points* (minPts) dan *Epsilon* (eps) yang digunakan. *Cluster* yang dibentuk didasarkan pada “kepadatan” (*density*) data. Konsep kepadatan pada metode DBSCAN akan menghasilkan tiga status dari setiap data, yaitu *noise*, inti (memenuhi Eps dan minPts), dan batas (*border*). Metode DBSCAN sangat cocok dalam mengelompokkan data-data dengan kerapatan tinggi, tidak memerlukan informasi tentang banyaknya kelompok yang akan dibentuk, dan hanya membutuhkan dua parameter yang sebagian besar tidak sensitif terhadap basis data [6]. Berbeda dengan DBSCAN, *K-Means* membagi data menjadi *k* kelompok, setiap kelompok memiliki nilai *centroid* masing-masing. Kemudian dihitung jarak *euclidean* data dengan *centroid* masing-masing kelompok. Data yang memiliki jarak terdekat dengan *centroid* akan masuk ke dalam kelompok yang sama. Proses tersebut dilakukan hingga seluruh data terbagi menjadi *k-cluster*. Meskipun kedua metode tersebut sama-sama menggunakan jarak *euclidean*, namun *K-Means* tidak memperhatikan *noise* data sedangkan DBSCAN dapat mendeteksi adanya *noise* yaitu data yang tidak tergabung dalam kelompok manapun.

Maka dari itu, penulis tertarik untuk membuat penelitian yang berjudul Komparasi DBSCAN dan *K-Means Clustering* pada Pengelompokan Status Desa di Jawa Tengah Tahun 2020 dengan tujuan mengklasifikasikan status desa berdasarkan kemiripan nilai IKL, IKS, dan IKE. Selain itu, penelitian ini bertujuan untuk mengetahui apakah status desa yang sama juga mengelompok pada *cluster* yang sama serta membandingkan dari kedua metode *clustering* tersebut manakah yang lebih baik dalam mengklasifikasikan status desa di Provinsi Jawa Tengah tahun 2020 berdasarkan nilai koefisien *Silhouette*.

2. TINJAUAN PUSTAKA

2.1 Indeks Desa Membangun

Indeks Desa Membangun (IDM) merupakan Indeks Komposit yang dibentuk berdasarkan tiga indeks, yaitu Indeks Ketahanan Sosial, Indeks Ketahanan Ekonomi dan Indeks Ketahanan Ekologi/Lingkungan. Perangkat indikator dalam IDM dikembangkan berdasarkan konsepsi bahwa untuk menuju Desa maju dan mandiri perlu dibuat suatu kerangka kerja pembangunan

yang berkelanjutan dalam rangka mencapai kesejahteraan kehidupan desa di segala aspek baik aspek sosial, ekonomi, maupun ekologi.

2.2 Status Desa

Status desa ditentukan berdasarkan nilai Indeks Desa Membangun (IDM). Setiap desa memiliki IDM yang dihitung berdasarkan rata-rata nilai IKL, IKS, dan IKE. Berdasarkan informasi Kementerian Desa, nilai rata-rata nasional IDM di Indonesia adalah 0,566. Adapun ambang batas dalam mengklasifikasikan status desa berdasarkan tabel berikut:

Tabel 2.1. Kategori Status Desa

No	Status Desa	IDM
1.	Desa Mandiri	$\geq 0,8115$
2.	Desa Maju	$<0,8115$ dan $>0,7072$
3.	Desa Berkembang	$<0,7072$ dan $>0,5989$
4.	Desa Tertinggal	$<0,5989$ dan $>0,4907$
5.	Desa Sangat Tertinggal	$\leq 0,4907$

Sumber: Kementerian Desa, 2019

Klasifikasi terhadap status desa tersebut bertujuan untuk penetapan status perkembangan dan rekomendasi terhadap intervensi kebijakan yang perlu dilakukan pada tahun 2019. Sehingga pendekatan dan intervensi yang diterapkan pada Status Desa Tertinggal, Desa Tertinggal, Desa Berkembang, Desa Maju, dan Desa Mandiri akan berbeda sesuai hasil updating data di Tahun 2018.

2.3 DBSCAN

Density-Based Spatial Clustering of Application with Noise (DBSCAN) merupakan metode pengelompokan yang didasarkan pada kepadatan data, yaitu banyaknya data (*minimum points*) yang berada dalam radius Eps (ϵ) dari setiap data [5]. Konsep kepadatan seperti ini menghasilkan tiga macam status dari setiap data, yaitu *noise* (titik yang tidak dapat dijangkau oleh *core* dan bukan merupakan *border*), *core* (titik pusat dalam *cluster* didasarkan pada densitas dimana ada sejumlah titik yang harus berada dalam Eps), dan *border* (titik yang menjadi batasan dalam kawasan titik pusat). DBSCAN akan melakukan *clustering* sesuai dengan parameter masukannya yaitu $epsilon$ dan $minpts$. Jumlah *cluster* yang dihasilkan DBSCAN akan bergantung pada kedua parameter tersebut. Jarak *Euclidean* adalah jarak yang digunakan dalam proses pengelompokan menggunakan metode DBSCAN dengan menghitung jarak antara suatu titik ke *centroid* (C) yang telah ditentukan secara acak sebelumnya. Secara matematis dapat dirumuskan sebagai berikut [3].

$$D(x_i, C_i) = \sqrt{\sum_{j=1}^q (x_{ij} - C_{ij})^2} \quad (2.1)$$

Keterangan :

$D(x_i, C_i)$ = Jarak antara titik x_i dengan titik C_i

x_{ij} = nilai titik i pada *cluster* ke- j

C_{ij} = nilai *centroid* i pada *cluster* ke- j

2.4 K-Means

K-Means merupakan salah satu metode *clustering* non hierarki yang mengelompokkan data ke dalam satu atau lebih *cluster*, berdasarkan kemiripan karakteristik tiap-tiap data [7]. Data dengan karakteristik yang berbeda dikelompokkan dalam kelompok yang lain. Metode *K-Means* memiliki beberapa keunggulan diantaranya yaitu proses penghitungan matematika yang sederhana, konvergensi yang cepat dan mudah diaplikasikan [11].

Pengelompokkan data dengan metode *K-Means* dilakukan dengan algoritma sebagai berikut [10].

1. Menentukan jumlah kelompok.
2. Data dialokasikan ke dalam kelompok secara acak.
3. Hitung pusat kelompok dari data yang ada di masing-masing kelompok. Lokasi *centroid* setiap kelompok diambil dari rata-rata semua nilai data dari setiap fiturnya. Jika M adalah jumlah data dalam kelompok, i adalah banyak kelompok, dan p adalah dimensi data, maka persamaan untuk menghitung *centroid* fitur ke- i digunakan persamaan (2.2).

$$K^i = \frac{1}{M} \sum_{j=1}^M X_j \quad (2.2)$$

Keterangan :

K^i = *centroid* ke- i

M = jumlah data dalam kelompok

X_j = titik ke- j

4. Alokasikan masing-masing data ke *centroid* terdekat. Menghitung jarak data ke pusat kelompok dengan jarak *Euclidean* yang dinyatakan dalam persamaan (2.3)

$$d = \sqrt{(x_1 - x_2)^2 + (c_1 - c_2)^2} \quad (2.3)$$

Keterangan :

d = jarak *Euclidean*

x_j = titik ke- j

c_j = *centroid* ke- j

5. Apabila masih ada perubahan nilai *centroid* di atas nilai ambang yang ditentukan, maka dilakuna penghitungan kembali pada langkah 3 dan seterusnya.

2.4 Koefisien *Silhouette*

Indeks *silhouette* menjadi salah satu ukuran dalam menguji validitas dari hasil pengelompokkan metode yang digunakan. Nilai koefisien *silhouette* dinyatakan dalam persamaan (2.4) [4].

$$S(h) = \frac{b(h) - a(h)}{(a(h), b(h))} \quad (2.4)$$

Dimana nilai $a(h)$ yaitu rata-rata jarak titik inti dengan semua titik pada kelompok yang sama, dan nilai $b(h)$ yaitu rata-rata jarak titik inti dengan semua titik pada kelompok yang berbeda menggunakan jarak *Euclidean*. Koefisien *Silhouette* menghasilkan rentang nilai -1 hingga 1, semakin baik pengelompokan data dalam satu *cluster*, maka hasilnya mendekati nilai 1 dan sebaliknya.

3. METODE PENELITIAN

3.1 Teknik Pengumpulan Data

Jenis data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Kementerian Desa Pembangunan Daerah Tertinggal dan Transmigrasi Republik Indonesia

tahun 2020. Variabel yang digunakan adalah Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL) yang didapatkan dari hasil *crawling web* Portal Data Desa Indonesia.

3.2 Tahapan Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah *clustering* menggunakan metode DBSCAN dan *K-Means* menggunakan *software* R. Tahap analisis yang dilakukan adalah sebagai berikut:

1. Melakukan pengumpulan dan *preprocessing* data.
2. Melakukan *clustering* status desa menggunakan metode DBSCAN.
3. Kemudian melakukan *clustering* status desa menggunakan metode *K-Means*.
4. Validasi *cluster* menggunakan koefisien *silhouette* pada kelompok yang telah dibentuk menggunakan metode DBSCAN dan *K-Means*.
5. Setelah itu dilakukan analisis pembahasan dari hasil pengelompokkan kedua metode tersebut.

4. HASIL DAN PEMBAHASAN

4.1 Ringkasan Data

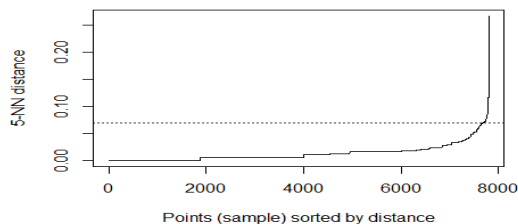
Tabel 4.1 Ringkasan Nilai IKS, IKE, IKL, dan IDM Berdasarkan Status Desa di Provinsi Jawa Tengah 2020

Status Desa	Rata-Rata			
	IKS	IKE	IKL	IDM
Desa Berkembang	0,7559	0,5797	0,6441	0,6598
Desa Maju	0,8117	0,7066	0,7397	0,7527
Desa Mandiri	0,8764	0,8488	0,8339	0,8530
Desa Sangat Tertinggal	0,6930	0,4958	0,4000	0,5296
Desa Tertinggal	0,6970	0,4649	0,5380	0,5664

Sumber data: *Crawling data website* Portal Data Desa Indonesia, diolah

Berdasarkan tabel diatas, terlihat bahwa untuk status desa mandiri rata-rata IKS desa sebesar 0.8764, IKE 0.8488, IKL 0.8339 dengan rata-rata IDM status desa mandiri sebesar 0.8530. Untuk status desa maju memiliki rata-rata IKS sebesar 0.8117, IKE dengan rata-rata sebesar 0.7066, IKL 0.7397 dan dengan rata-rata IDM sebesar 0.7527. Status desa berkembang dari 5.564 desa memiliki rata-rata IKS sebesar 0.7559, rata-rata IKE sebesar 0.5797, IKL sebesar 0.6441 dan rata-rata IDM sebesar 0.6598. Desa dengan status tertinggal dari 262 desa memiliki rata-rata IKS sebesar 0.6970, rata-rata IKE sebesar 0.4649, rata-rata IKL 0.5480 dan dengan rata-rata IDM sebesar 0.5664. Sedangkan untuk desa sangat tertinggal memiliki rata-rata IKS 0.6930, rata-rata IKE 0.4958, IKL 0.4000 dan rata-rata IDM dari kedua desa dengan status sangat tertinggal sebesar 0.5296.

4.2 *Clustering* Status Desa Menggunakan Metode DBSCAN



Gambar 4.1 KNN Plot untuk Dataset Status Desa di Provinsi Jawa Tengah 2020

Berdasarkan gambar 4.1, dapat dilihat bahwa secara visual dengan menerapkan *5-NN distance* (5 jarak ketetanggaan terdekat), nilai optimal dari *epsilon (Eps)* berada pada rentang 0.05 sampai 0.1. Setelah dilakukan beberapa kali percobaan, peneliti memutuskan menggunakan nilai *epsilon* 0.072 dan *minPts* atau minimal titik dalam satu *cluster* adalah 2 dengan pertimbangan bahwa di Provinsi Jawa Tengah pada tahun 2020, jumlah desa minimum dari 5 jenis status desa adalah 2 desa yaitu pada desa dengan status sangat tertinggal. Setelah dilakukan analisis metode DBSCAN dengan menggunakan *software* R, didapatkan hasil pengelompokan jumlah desa di setiap status desa adalah sebagai berikut.

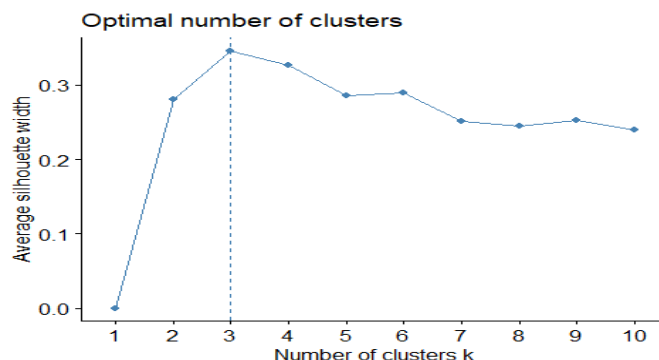
Tabel 4.2 Hasil *Clustering* Status Desa di Provinsi Jawa Tengah Tahun 2020 Menggunakan Metode *DBSCAN*

Cluster	Berkembang	Maju	Mandiri	Sangat Tertinggal	Tertinggal
0	4	1	3	2	3
1	5556	1836	137	0	255
2	2	0	0	0	0
3	2	0	0	0	0
4	0	0	0	0	4

Sumber: *Crawling data website* Portal Data Desa Indonesia, diolah

Berdasarkan hasil tabel 4.2, dapat diketahui bahwa *cluster* yang terbentuk sebanyak 4 *cluster* dan disertai adanya *noise* data (*cluster* 0) sebanyak 13 desa. Artinya dengan menggunakan metode DBSCAN terdapat 13 desa yang tidak tergabung dalam *cluster* manapun. Selain itu, dapat kita lihat pula bahwa *cluster* yang memuat desa paling banyak adalah *cluster* 1 yang mana sebagian besar desa-desa di Provinsi Jawa Tengah tahun 2020 pada *cluster* tersebut tergolong sebagai desa dengan status berkembang.

4.3 *Clustering* Status Desa Menggunakan Metode *K-Means*



Gambar 4.2 Jumlah *Cluster Optimum* Berdasarkan Nilai Koefisien *Silhouette*

Pada gambar 4.2, terlihat bahwa nilai koefisien *silhouette* yang optimal berada pada $k = 3$. Oleh karena itu, pada penelitian ini dilakukan analisis dengan metode *K-Means* yang menggunakan jumlah *cluster*/kelompok sebanyak 3 (tiga). Berikut hasil analisisnya yang disajikan dalam Tabel 4.3.

Tabel 4.3 Hasil *Clustering* Status Desa di Provinsi Jawa Tengah Tahun 2020 dengan Menggunakan Metode *K-Means*

Cluster	Berkembang	Maju	Mandiri	Sangat Tertinggal	Tertinggal
1	3969 (92.20)	82 (1.90)	0 (0.00)	1 (0.02)	252 (5.85)
2	1036 (49.22)	990 (47.03)	67 (3.18)	1 (0.05)	9 (0.43)
3	559 (39.90)	765 (54.60)	73 (5.21)	0 (0.00)	1 (0.07)

Sumber: *Crawling* data dari *website* Portal Data Desa Indonesia, diolah

Berdasarkan hasil tabel 4.3. dapat diketahui bahwa *cluster* yang terbentuk sebanyak 3. Selain itu, dapat kita lihat pula bahwa dengan menggunakan metode *K-Means cluster* yang memuat jumlah desa paling banyak di Provinsi Jawa Tengah Tahun 2020 adalah *cluster* 1 dengan proporsi terbesar yaitu 92,20 persen berada pada status berkembang.

4.4 Perbandingan Nilai Koefisien *Silhouette* Metode DBSCAN dan *K-Means*

Tabel 4.4 Nilai Koefisien *Silhouette* dengan Menggunakan Metode DBSCAN

Cluster	Jumlah Desa	Rata-rata Nilai Koefisien <i>Silhouette</i>
0	13	-0,49
1	7784	0,42
2	2	0.68
3	2	0.93
4	4	0.83

Sumber: *Crawling data website* Portal Data Desa Indonesia, diolah

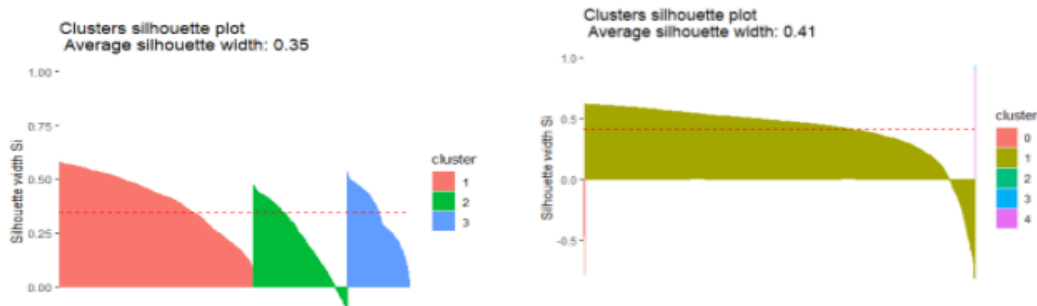
Hasil perhitungan menunjukkan bahwa desa-desa di Provinsi Jawa Tengah pada tahun 2020 terkonsentrasi pada *cluster* 1 dengan rata-rata nilai koefisien *silhouette* sebesar 0.42. Akan tetapi, jika dilihat secara keseluruhan rata-rata nilai koefisien *silhouette* diatas, *cluster* yang paling baik dalam mengelompokkan desa-desa tersebut karena nilainya mendekati 1 adalah *cluster* 3 dan *cluster* 4.

Tabel 4.5 Nilai Koefisien *Silhouette* dengan Menggunakan Metode *K-Means*

<i>Cluster</i>	Jumlah Desa	Rata-rata Nilai Koefisien <i>Silhouette</i>
1	4317	0,41
2	2090	0,23
3	1398	0,34

Sumber: *Crawling data website* Portal Data Desa Indonesia, diolah

Di sisi lain, jika menggunakan metode *K-Means*, hasilnya menunjukkan bahwa desa-desa di Provinsi Jawa Tengah pada tahun 2020 terkonsentrasi pada *cluster* 1 dengan rata-rata nilai koefisien *silhouette* sebesar 0.41 dengan jumlah desa sebanyak 4317 dan sekaligus menjadi *cluster* paling baik dibandingkan *cluster* 2 dan 3 dalam mengelompokkan desa-desa tersebut.



Gambar 4.3 Rata-Rata Nilai *Silhouette* (a) Metode *DBSCAN* (b) Metode *K-Means*

Berdasarkan gambar 4.3 secara visual yang ditunjukkan oleh garis putus-putus menunjukkan bahwa rata-rata nilai koefisien *silhouette* metode *DBSCAN* lebih besar dibandingkan dengan metode *K-Means*. Hal ini berarti bahwa metode *DBSCAN* lebih baik digunakan daripada metode *K-Means* dalam mengelompokkan desa-desa di Provinsi Jawa Tengah berdasarkan nilai *IKE*, *IKL*, dan *IKS* pada tahun 2020.

Selain itu, metode *DBSCAN* juga dapat mendeteksi desa-desa pada kelompok *noise* (desa yang tidak termasuk pada kelompok manapun) yang mana jumlahnya ada sebanyak 13 desa. Dari hasil pengelompokan yang dilakukan oleh metode *DBSCAN* dapat dilihat juga bahwa tiap-tiap *cluster* mengandung desa dengan status yang berbeda-beda dibandingkan dengan pengelompokan yang dilakukan oleh Kementerian Desa. Misalnya, pada *cluster* 1 terdapat 92.20 persen desa di Jawa Tengah tahun 2020 dengan status desa berkembang, dan sisanya terdiri dari status desa maju, tertinggal, dan sangat tertinggal.

Hasil pengelompokan yang berbeda tersebut dapat disebabkan karena adanya perbedaan dalam proses pembentukan kelompok yang mana metode *DBSCAN* berdasarkan kemiripan karakteristik dan kepadatan nya sedangkan pengelompokan yang dilakukan oleh Kementerian

Desa hanya berdasarkan rata-rata dari IKE, IKL, dan IKS dengan ambang batas yang telah ditentukan sebelumnya. Oleh karena itu, hal ini tentu mengindikasikan bahwa di Jawa Tengah tahun 2020, status desa yang dikelompokkan oleh Kementerian Desa belum mempertimbangkan adanya kemiripan karakteristik dan kepadatan dari ketiga indeks tersebut.

5. KESIMPULAN

Berdasarkan hasil pembahasan yang telah disampaikan diatas, maka dapat disimpulkan sebagai berikut:

1. Data IKS, IKE, dan IKL di Provinsi Jawa Tengah berhasil di *clustering* dan diterapkan ke dalam metode *DBSCAN* dan *K-Means clustering*. Penentuan nilai *Epsilon* dan *minPts* pada metode *DBSCAN* akan berpengaruh langsung terhadap jumlah cluster yang dihasilkan.
2. Hasil pengelompokan desa-desa di Provinsi Jawa Tengah tahun 2020 yang menggunakan metode *DBSCAN* dan *K-Means* pada data IKE, IKS, dan IKL berbeda dengan recap hasil status desa pada PDDI Kementerian Desa, hal ini dikarenakan metode *clustering* bekerja dengan mengelompokkan data yang mirip dalam hirarki yang sama dan yang tidak mirip di hirarki yang agak jauh serta hasil ini juga mengindikasikan bahwa status yang dikelompokkan oleh Kementerian Desa dinilai belum mempertimbangkan adanya kemiripan karakteristik pada ketiga indeks yang dihasilkan tersebut.
3. Berdasarkan rata-rata nilai koefisien *silhouette*, metode *DBSCAN* lebih baik digunakan daripada metode *K-Means* dalam mengelompokkan desa-desa di Provinsi Jawa Tengah berdasarkan nilai IKE, IKL, dan IKS pada tahun 2020. Hal tersebut terlihat dari nilai koefisien *silhouette* yang dihasilkan metode *DBSCAN* lebih mendekati 1 dibandingkan metode *K-Means*.

DAFTAR PUSTAKA

- [1] Apsari, W. 2019. *Mendes PDDT Optimis Target RPJMN 2020-2024 Dapat Direalisasikan*. Monitor, Jakarta. <https://monitor.co.id/2019/11/27/mendes-pdtt-optimis-target-rpjmn-2020-2024-dapat-direalisasikan>. [11 Oktober 2020]
- [2] BPS. 2020. *Berita Resmi Statistik: Pertumbuhan Ekonomi Indonesia Triwulan II-2020*. Publikasi BPS, Jakarta. <https://www.bps.go.id/pressrelease/2020/08/05/1737/-ekonomi-indonesia-triwulan-ii-2020-turun-5-32-persen.html>. [13 September 2020]
- [3] Budiman S.A.D., Safitri D. & Ispriyanti D., 2016. Perbandingan Metode K-Means dan Metode DBSCAN pada Pengelompokan Rumah Kost Mahasiswa di Kelurahan Tembalang Semarang. *Jurnal Gaussian*, Vol. 5, No. 4, 757-762.
- [4] Kristianto A., Sedyono E. & Hartomo Kristoko D., 2020. Implementation DBSCAN algorithm to clustering satellite surface temperature data in Indonesia. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, Vol. 6, No. 2, 109-118.
- [5] Nooraeni Rani, 2016. Metode Cluster Menggunakan Kombinasi Algoritma Cluster K-Prototype Dan Algoritma Genetika Untuk Data Bertipe Campuran. *Jurnal Aplikasi Statistika & Komputasi Statistik*, Vol. 7, No. 2, 81-98.
- [6] Novi S., 2020. Clustering Status Desa Menggunakan Metode DBSCAN (Density-Based Spatial Clustering of Application With Noise). *Thesis (Skripsi)*. Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Riau.

- [7] Pramana S., Yuniarto B., Mariyah S., Santoso I. & Nooraeni R., 2018. *Data Mining dengan R : Konsep Serta Implementasi*. In Media., Bogor.
- [8] Pusat Data Desa Indonesia. 2020. *Indeks Desa Membangun*. Kementerian Desa Pembangunan Daerah Tertinggal dan Transmigrasi, Jakarta. <https://pddi.kemendes.go.id/idm>. [13 September 2020]
- [9] Satuan Tugas Covid-19. 2020. *Peta Sebaran Covid-19*. <https://covid19.go.id/peta-sebaran>. [22 September 2020]
- [10] Talakua M.W., Leleury Z.A. & Taluta A.W., 2017. Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014. *Register: A. BAREKENG: Jurnal Ilmu Matematika dan Terapan*, Vol. 11, No. 2, 119-128.
- [11] Yuan C. & Yang H., 2019. Research on K-value selection method of K-means clustering algorithm. *J—Multidisciplinary Scientific Journal*, Vol. 2, No. 2, 226-235.