

**KLASTERISASI METODE *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN)* DAN *K-MEANS* DALAM
PENGELOMPOKAN PROVINSI BERDASARKAN FAKTOR PENYEBAB
STUNTING DI INDONESIA**

(Skripsi)

Oleh

NADIA ATHIYAH PUTRI



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

ABSTRACT

CLUSTERING METHOD OF DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN) AND K-MEANS IN CLUSTERING PROVINCES BASED ON FACTORS THAT CAUSE STUNTING IN INDONESIA

By

NADIA ATHIYAH PUTRI

Clustering is a technique used to group data with similar characteristics within each cluster. There are several clustering methods available, such as K-Means and DBSCAN. K-Means is a clustering algorithm that partitions data into several clusters, while DBSCAN builds clusters based on connected densities. This research will use K-Means and DBSCAN clustering methods to perform clustering with stunting causal factor data and simulated data as a consideration of the performance of both methods. The results obtained on the data of the factors that cause stunting, the K-Means method has better performance based on the SI value of 0.5253 and the CH value of 26.0940. While in simulation data where the data is normally distributed and does not contain outliers, the DBSCAN method has better performance based on the SI value of 0.6495 and the CH value of 328.0302.

Kata kunci : Clustering, K-Means, DBSCAN, Stunting.

ABSTRAK

KLASTERISASI METODE *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE* (DBSCAN) DAN *K-MEANS* DALAM PENGELOMPOKAN PROVINSI BERDASARKAN FAKTOR PENYEBAB *STUNTING* DI INDONESIA

Oleh

NADIA ATHIYAH PUTRI

Clustering adalah teknik yang digunakan untuk mengelompokkan data dengan karakteristik yang sama dalam setiap *cluster*. Ada beberapa metode *clustering* yang tersedia, seperti *K-Means* dan DBSCAN. *K-Means* adalah algoritma *clustering* yang mempartisi data ke dalam beberapa *cluster*, sedangkan DBSCAN membangun *cluster* berdasarkan densitas yang terhubung. Penelitian ini akan menggunakan metode *clustering K-Means* dan DBSCAN untuk melakukan *clustering* dengan data faktor penyebab *stunting* dan data simulasi sebagai pertimbangan performa dari kedua metode. Hasil yang diperoleh pada data faktor penyebab *stunting* metode *K-Means* memiliki performa yang lebih baik berdasarkan nilai SI sebesar 0.5253 dan nilai CH sebesar 26.0940. Sedangkan pada data simulasi dimana data tersebut berdistribusi normal dan tidak mengandung pencilan metode DBSCAN memiliki performa yang lebih baik berdasarkan nilai SI sebesar 0,6495 dan nilai CH sebesar 328,0302.

Kata kunci : *Clustering, K-Means, DBSCAN, Stunting.*

KLASTERISASI METODE *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN)* DAN *K-MEANS* DALAM PENGELOMPOKAN PROVINSI BERDASARKAN FAKTOR PENYEBAB *STUNTING* DI INDONESIA

Oleh

**NADIA ATHIYAH PUTRI
2017031006**

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

Judul Skripsi

KLASTERISASI METODE *DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN)* DAN *K-MEANS* DALAM PENGELOMPOKAN PROVINSI BERDASARKAN FAKTOR PENYEBAB *STUNTING* DI INDONESIA

Nama Mahasiswa

Nadia Athiyah Putri

Nomor Pokok Mahasiswa

: 2017031006

Program Studi

: Matematika

Fakultas

: Matematika dan Ilmu Pengetahuan Alam



MENYETUJUI

1. Komisi Pembimbing



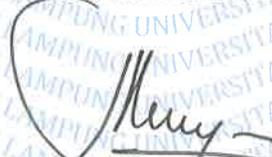
Dr. Khoirin Nisa, S.Si., M.Si.
NIP. 19740726 200003 2 001



Dra. Dorrah Aziz, M.Si.
NIP. 19610128 198811 2 001

Mengetahui,

2. Ketua Jurusan Matematika



Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Khoirin Nisa, S.Si., M.Si.**



Sekretaris : **Dra. Dorrah Aziz, M.Si.**

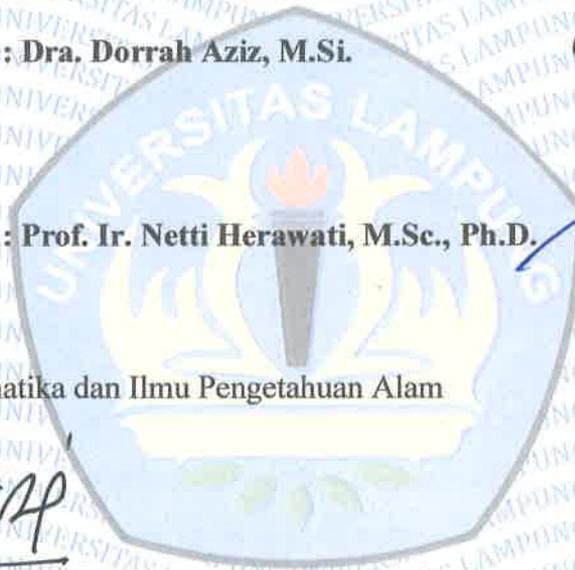


Penguji
Bukan Pembimbing : **Prof. Ir. Netti Herawati, M.Sc., Ph.D.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711001 200501 1 002



Tanggal Lulus Ujian Skripsi : **01 Juli 2024**

PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Nadia Athiyah Putri
Nomor Pokok Mahasiswa : 2017031006
Jurusan : Matematika
Judul Skripsi : **Klasterisasi Metode *Density-Based Spatial Clustering Of Application With Noise* (DBSCAN) dan *K-Means* dalam Pengelompokan Provinsi Berdasarkan Faktor Penyebab *Stunting* di Indonesia**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku

Bandar Lampung, 01 Juli 2024
Penulis,



Nadia Athiyah Putri
2017031006

RIWAYAT HIDUP

Penulis bernama Nadia Athiyah Putri, anak tunggal dari Bapak Agustomi, S.Pd. dan Ibu Sriyani yang lahir di Kotabumi pada hari Kamis, 22 Agustus 2002. Penulis menyelesaikan pendidikan sekolah dasar di SD Negeri 04 Tanjung Aman pada tahun 2008 s.d 2014, sekolah menengah pertama di SMP Negeri 01 Kotabumi pada tahun 2014 s.d 2017, dan sekolah menengah atas di SMA Negeri 03 Kotabumi pada tahun 2017 s.d 2020.

Pada tahun 2020 penulis diterima sebagai mahasiswa S1 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur SNMPTN. Selama aktif menjadi mahasiswa, penulis ikut serta dalam organisasi Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) sebagai anggota Bidang Minat dan Bakat pada tahun 2021 dan Badan Eksekutif Mahasiswa (BEM) FMIPA pada tahun 2022.

Pada tahun 2023, sebagai bentuk penerapan bidang ilmu di dunia kerja, penulis melaksanakan Kerja Praktek (KP) di Badan Pusat Statistik (BPS) Kabupaten Lampung Utara dan sebagai bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Rantau Jaya Baru, Kecamatan Putra Rumbia, Kabupaten Lampung Tengah.

KATA INSPIRASI

*“Sebaik-baiknya doa adalah doa yang diridhai oleh orang tua”
(Penulis).*

*“Allah tidak membebani seseorang melainkan sesuai dengan
kesanggupannya”
(QS. Al-Baqarah: 286).*

*“Long story short, it was a bad time. Long story short, I survived”
(Taylor Swift)*

*“If you never bleed, you never grow
(Taylor Swift)*

*“Bunga yang mekar dalam kesulitan adalah yang paling langka dan indah
dari semua”
(Kisar from Mulan)*

*"Berikan hal baik untuk sekitarmu, maka semesta akan mendekatkan hal
baik kepadamu"
(Penulis)*

PERSEMBAHAN

Alhamdulillahrabbi'lamin
Puji Syukur tiada hentinya terhanturkan kepada Allah SWT
Kupersembahkan karya ini kepada:

Ayah, dan Bunda

Orang tuaku tercinta, Bapak Agustomi, S.Pd.Gr., dan Ibu Sriyani. Terima kasih sudah menjadi alasan penulis untuk terus berjuang menyelesaikan skripsi ini. Terima kasih selalu memeluk penulis saat gagal, dan mengapresiasi penulis untuk seluruh pencapaian kecil yang telah penulis raih. Semoga ayah dan bunda bangga dengan usahaku ini.

Dosen

Dosen-dosen pembimbing dan pembahas yang sangat berjasa dalam membimbing serta memberikan saran yang membangun dalam proses penyelesaian skripsi ini.

Diri Sendiri

Terima kasih sudah percaya dengan diriku sendiri. Perjalanan dan proses ini tidak mudah, untungnya ku bisa rasa hal-hal baik yang datangnya belakangan. *I gave my blood, sweat, and tears for this.*

Sahabat-sahabatku

Para sahabat terkasih yang selalu mendukung, membantu, serta memberikan warna dalam hidupku. *Thank you for being my unpaid therapist.*

Almamater kebanggaan, Universitas Lampung.

SANWACANA

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasterisasi Metode *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) dan *K-Means* dalam Pengelompokan Provinsi Berdasarkan Faktor Penyebab *Stunting* di Indonesia” secara maksimal. Berhasilnya penyelesaian skripsi ini tak lepas dari bimbingan, dorongan serta arahan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku Dosen Pembimbing I yang senantiasa memberikan dukungan kepada penulis dalam bentuk ilmu, bimbingan, serta saran dalam menyelesaikan skripsi ini.
2. Ibu Dra. Dorrah Aziz, M.Si., selaku Dosen Pembimbing II yang juga senantiasa memberikan dukungan kepada penulis dalam bentuk ilmu, bimbingan, dan saran dalam menyelesaikan skripsi ini.
3. Ibu Prof. Ir. Netti Herawati, M.Sc., Ph.D., selaku Dosen Penguji Skripsi yang telah memberikan kritik, evaluasi, serta saran dalam perbaikan skripsi penulis.
4. Ibu Prof. Dra. Wamiliana, MA, Ph.D., selaku dosen pembimbing akademik.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika.
6. Seluruh Dosen, Staf, dan Karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Ayah dan Bunda yang senantiasa memberikan dukungan, doa, cinta, nasihat, motivasi, kepercayaan, dan rumah yang nyaman kepada penulis sehingga penulis bisa terus berjuang menyelesaikan skripsi ini.

8. Sahabat penulis, Devanisa Norma Deviana, yang senantiasa memberikan motivasi, dukungan, serta meluangkan waktunya dalam membantu penulis guna menyelesaikan skripsi ini.
9. Sahabat terkasih, Aprilia, Nazmah, Wahyu, Salsa, Dzakiyah, Asha, dan Caca yang selalu memberikan dukungan berupa hiburan, semangat, motivasi, dan bantuan kepada penulis. *Thank you for being my unpaid therapist.*
10. Teman-teman penulis, Mawar, Oca, Rini, Silvia yang selalu memberikan dukungan serta motivasi dan bersedia berbagi pengalaman kepada penulis.
11. Partner penulis, yang senantiasa memberikan motivasi, energi positif, dan hiburan serta bersedia direpotkan oleh penulis.
12. Anak kecil lucu, Dmitriev Abraham Hariyanto, yang tingkah lakunya senantiasa menjadi hiburan bagi penulis dalam menyelesaikan skripsi ini. Abadam, terima kasih sudah hadir di dunia ini.
13. Semua pihak yang membantu dalam penyelesaian skripsi ini.

Penulis menyadari sepenuhnya bahwa dalam penulisan skripsi ini masih banyak kekurangan. Oleh karena itu, penulis memohon maaf atas segala kesalahan dan kekurangan yang terdapat dalam skripsi ini. Akhir kata, semoga Tuhan Yang Maha Esa membalas semua kebaikan seluruh pihak yang membantu dalam proses penyelesaian skripsi ini. Semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Bandar Lampung, 01 Juli 2024

Penulis

Nadia Athiyah Putri

2017031006

DAFTAR ISI

	Halaman
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
1. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian.....	5
2. TINJAUAN PUSTAKA	6
2.1 <i>Data Mining</i>	6
2.2 <i>Clustering</i>	8
2.3 Jarak Euclidean	8
2.4 <i>Density-Based Spatial Clustering of Application with Noise (DBSCAN)</i> ...	9
2.5 <i>K-Means</i>	10
2.6 <i>Silhouette Coefficient</i>	12
2.7 <i>Calinski-Harabasz (CH)</i>	13
2.8 Metode <i>Elbow</i>	14
2.9 <i>Stunting</i>	15
3. METODOLOGI PENELITIAN	16
3.1 Waktu dan Tempat Penelitian	16
3.2 Data Penelitian	16
3.3 Metode Penelitian	17
4. HASIL DAN PEMBAHASAN	19
4.1 Analisis Deskriptif	19
4.2 <i>Preprocessing</i> Data	23
4.3 Proses Klusterisasi Data	25
4.3.1 Klusterisasi Menggunakan Metode DBSCAN	25
4.3.2 Klusterisasi Menggunakan Metode <i>K-Means</i>	36
4.4 Perbandingan Indeks Validitas Metode DBSCAN dan <i>K-Means</i>	47
5. KESIMPULAN	49

DAFTAR PUSTAKA	51
LAMPIRAN	55

DAFTAR TABEL

Tabel	Halaman
1. Pengelompokan berdasarkan <i>Silhouette coefficient</i>	13
2. Statistika deskriptif untuk data faktor penyebab <i>stunting</i>	19
3. Statistika deskriptif untuk data simulasi	21
4. Data faktor penyebab <i>stunting</i> hasil <i>preprocessing</i>	24
5. Data simulasi hasil <i>preprocessing</i>	25
6. Percobaan nilai parameter DBSCAN dengan beberapa nilai epsilon dan <i>min-points</i> pada data faktor penyebab <i>stunting</i>	26
7. Percobaan nilai parameter DBSCAN dengan beberapa nilai epsilon dan <i>min-points</i> pada data simulasi	29
8. Hasil clustering data faktor penyebab <i>stunting</i> dengan metode DBSCAN ...	31
9. Hasil <i>clustering</i> data simulasi dengan metode DBSCAN.....	32
10. Anggota setiap <i>cluster</i> pada data faktor penyebab <i>stunting</i> dengan metode DBSCAN	35
11. Nilai rata-rata setiap variabel data faktor penyebab <i>stunting</i> pada <i>cluster</i> yang terbentuk.....	35
12. Nilai rata-rata setiap variabel data simulasi pada <i>cluster</i> yang terbentuk	36
13. Percobaan jumlah <i>cluster</i> metode <i>K-means</i> beserta nilai <i>Silhouette coefficient</i> dan <i>Callinski-Harabasz</i> pada data faktor penyebab <i>stunting</i>	39
14. Percobaan jumlah <i>cluster</i> metode <i>K-means</i> beserta nilai <i>Silhouette coefficient</i> dan <i>Callinski-Harabasz</i> pada data	39
15. Hasil <i>clustering</i> data faktor penyebab <i>stunting</i> metode <i>K-means</i>	44

16. Hasil <i>clustering</i> data simulasi metode <i>K-means</i>	45
17. Anggota setiap <i>cluster</i> data faktor penyebab <i>stunting</i> pada metode <i>K-Means</i>	46
18. Nilai rata-rata setiap variabel data faktor penyebab <i>stunting</i> pada cluster yang terbentuk.....	46
19. Nilai rata-rata setiap variabel data simulasi pada cluster yang terbentuk	47
20. Nilai index validitas metode DBSCAN dan <i>K-means</i> pada data faktor penyebab <i>stunting</i>	47
21. Nilai index validitas metode DBSCAN dan <i>K-means</i> pada data simulasi.....	48

DAFTAR GAMBAR

Gambar	Halaman
1. Parameter Metode DBSCAN	10
2. <i>Box plot</i> data faktor penyebab <i>stunting</i>	20
3. <i>Box plot</i> data simulasi	22
4. Nilai <i>Silhouette coefficient</i> tertinggi disetiap <i>cluster</i> pada data faktor penyebab <i>stunting</i> metode DBSCAN	27
5. Nilai <i>Callinski-Harabsz</i> tertinggi disetiap <i>cluster</i> pada data faktor penyebab <i>stunting</i> metode DBSCAN	28
6. Nilai <i>Silhouette coefficient</i> tertinggi disetiap <i>cluster</i> pada data simulasi metode DBSCAN.....	30
7. Nilai <i>Callinski-Harabsz</i> tertinggi disetiap <i>cluster</i> pada data simulasi <i>stunting</i> metode DBSCAN	30
8. <i>Scatter plot clustering</i> data faktor penyebab <i>stunting</i> dengan metode DBSCAN	34
9. <i>Scatter plot clustering</i> data simulasi dengan metode DBSCAN.....	34
10. Metode <i>Elbow</i> data faktor penyebab <i>stunting</i>	37
11. Metode <i>Elbow</i> data simulasi	38
12. Nilai <i>Silhouette coefficient</i> setiap <i>cluster</i> pada data faktor penyebab <i>stunting</i> dengan metode <i>K-means</i>	40
13. Nilai <i>Callinski-Harabsz</i> setiap <i>cluster</i> pada <i>cluster</i> pada data faktor penyebab <i>stunting</i> denganmetode <i>K-means</i>	41
14. Nilai <i>Silhouette coefficient</i> setiap <i>cluster</i> pada data simulasi dengan metode <i>K-means</i>	41

15. Nilai <i>Callinski-Harabsz</i> setiap <i>cluster</i> pada <i>cluster</i> pada data simulasi dengan metode <i>K-means</i>	42
16. <i>Scatter plot clustering</i> data faktor penyebab <i>stunting</i> dengan metode <i>K-means</i>	43
17. <i>Scatter plot clustering</i> data simulasi dengan metode <i>K-means</i>	44

1. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Data *mining* merupakan proses yang menggunakan kecerdasan buatan (*Artificial Intelligence*), statistik, dan matematika dalam memperoleh pengetahuan dan informasi penting yang tersembunyi dari data (Jollyta, dkk., 2020). Dalam data *mining* terdapat beberapa tipe pemodelan data, yaitu *association*, *classification*, *clustering*, *forecasting*, *regression*, *sequence discovery*, dan *visualization*.

Clustering merupakan salah satu pemodelan data dalam data *mining* yang metode pembelajarannya tak terawasi (*unsupervised learning*) (Muflikhah, dkk., 2018).

Clustering digunakan untuk mengelompokkan suatu data yang memiliki karakteristik yang sama antara satu data dengan data yang lain dalam tiap *cluster*-nya. Terdapat tiga pendekatan dalam *clustering* yaitu *hierarchical clustering*, *density clustering*, dan *partitional clustering*. Pada pendekatan *hierarchical clustering*, terdapat beberapa metode yang seringkali digunakan yaitu *Divisive Analysis* (DIANA), dan *agglomerative hierarchical clustering*. Pada pendekatan *density clustering* terdapat beberapa metode salah satunya *Density-Based Spatial Clustering of Application with Noise* (DBSCAN). Sedangkan pada pendekatan *partitional clustering* terdapat beberapa metode seperti *K-Means*, dan *K-Medoids*.

K-Means merupakan salah satu algoritma *clustering* yang berbasis partisi yang berusaha mempartisi data yang ada ke dalam *cluster-cluster* (Mubarak & Gusmi, 2023). *K-Means* merupakan metode *clustering* yang paling terkenal karena mampu mengelompokkan berbagai objek ke masing-masing *cluster* dengan cepat dan sederhana. Selain itu *K-means* juga mudah diimplementasikan sehingga

banyak digunakan diberbagai bidang. Namun dalam penerapannya, metode ini membutuhkan jumlah *cluster* yang ditentukan terlebih dahulu sebagai k , dan kemudian membagi n objek ke dalam k *cluster*. Hal ini menjadi salah satu kelemahan dalam metode *K-means*.

Menurut Devi, dkk. (2015), DBSCAN adalah metode *clustering* yang membangun daerah pengelompokan berdasarkan kepadatan yang terkoneksi (*density-connected*) dimana pada setiap *cluster* akan memuat sejumlah minimum data dan mengidentifikasi data yang tidak masuk ke dalam *cluster* sebagai pencilan. Dengan demikian, kepadatan yang terkoneksi dalam metode DBSCAN memungkinkan pengelompokan data berdasarkan kepadatan yang terhubung secara langsung atau tidak langsung, tanpa memerlukan jumlah *cluster* yang telah ditentukan sebelumnya. DBSCAN akan secara otomatis membentuk *cluster* berdasarkan tingkat kemiripan suatu data dengan data yang lainnya sehingga akan menghasilkan suatu *cluster* yang memiliki karakteristik yang sama di tiap *clusternya*.

Merujuk pada penelitian sebelumnya Li, *et al.* (2023), yang menerapkan dua metode pengelompokan yaitu metode DBSCAN dan *K-means*. Kesimpulan dari penelitian ini adalah metode yang diusulkan menggunakan algoritma DBSCAN dan *K-means clustering* berhasil dalam melakukan penilaian risiko banjir perkotaan di Fuzhou, China. Metode ini mampu mengidentifikasi daerah dengan risiko banjir tinggi berdasarkan faktor-faktor seperti elevasi, kemiringan, permukaan yang tidak tembus air, dan koefisien aliran. Penelitian ini juga merekomendasikan langkah-langkah pencegahan risiko banjir yang sesuai untuk daerah dengan risiko tinggi yang diidentifikasi.

Penelitian lainnya seperti yang dilakukan Kristianto (2021), yang menganalisis performa metode DBSCAN dan *K-means* dalam *clustering* minat penggunaan transportasi umum. Penelitian ini menguji beberapa jumlah *cluster* dimana di setiap *cluster* nilai *silhouette index* dari metode DBSCAN cenderung mendekati

nilai 1 dan selalu lebih besar dibandingkan metode *K-means*. Kesimpulan yang diperoleh yaitu metode DBSCAN memiliki performa yang lebih baik dalam mengelompokkan minat penggunaan transportasi umum.

Penelitian yang dilakukan Adha, dkk. (2021), yang juga membandingkan metode DBSCAN dan *K-means* dalam mengelompokkan kasus covid-19 di dunia. Hasil dari penelitian ini adalah metode *K-means* lebih baik daripada metode DBSCAN dengan nilai *Silhouette coefficient* 0,6902 dan menghasilkan 8 *cluster*.

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa dalam melakukan perbandingan dengan metode DBSCAN dan *K-means* dalam mengelompokkan suatu data, hasil mengenai metode yang terbaik bergantung dengan hasil uji validitas yang dihasilkan. Selain itu, metode DBSCAN dan *K-means* juga dapat mengelompokkan berbagai jenis data termasuk data mengenai kesehatan sehingga dapat mempermudah dalam penyelesaian permasalahan tertentu.

Masalah kesehatan tak luput dari kehidupan setiap makhluk hidup mulai dari manusia, hewan, bahkan tumbuhan. Hal ini tidak menutup kemungkinan berbagai masalah kesehatan sudah dialami seseorang sejak balita. Salah satu contoh masalah kesehatan yang sering dialami balita di seluruh dunia adalah *stunting*. Masalah *stunting* merupakan masalah yang krusial khususnya di negara-negara miskin dan berkembang termasuk Indonesia. *Stunting* merupakan suatu kondisi dimana terjadi gagal tumbuh pada anak balita (bawah lima tahun) yang disebabkan oleh banyak faktor sehingga terjadi kelainan berupa anak terlalu pendek untuk usianya (Matdoan, dkk., 2022). *Stunting* diidentifikasi dengan menilai panjang atau tinggi badan anak (panjang badan berbaring untuk anak kurang dari 2 tahun dan tinggi berdiri untuk anak usia 2 tahun atau lebih) dan menafsirkan pengukuran dengan membandingkannya dengan serangkaian nilai standar yang dapat diterima (Onis & Francesco, 2016).

Kementerian Kesehatan Republik Indonesia mengumumkan hasil Survei Status Gizi Indonesia (SSGI) dimana prevalensi *stunting* di Indonesia pada tahun 2022 sebesar 21,6%. Hal ini masih melampaui standar *World Health Organization* (WHO) yang menetapkan prevalensi *stunting* harus diangka kurang dari 20% (Nuryuliyani, 2023). Untuk mendukung target pemerintah Indonesia dalam penurunan prevalensi *stunting* pada tahun 2024 yaitu sebesar 14%, diperlukan langkah-langkah yang tepat seperti mengelompokkan provinsi di Indonesia berdasarkan pemberian ASI eksklusif, rumah tangga dengan akses sanitasi layak, proses kelahiran dengan bantuan tenaga kesehatan terlatih, dan proses kelahiran di fasilitas kesehatan, sehingga diharapkan mampu memberikan dampak positif terhadap penurunan prevalensi *stunting* di Indonesia.

Berdasarkan permasalahan di atas, penulis tertarik untuk menulis penelitian dengan judul Klasterisasi Metode *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) dan *K-Means* dalam Pengelompokan Provinsi Berdasarkan Faktor Penyebab *Stunting* di Indonesia.

1.2 Tujuan Penelitian

Berdasarkan latar belakang dan masalah di atas, maka tujuan dari penelitian ini yaitu untuk mengetahui metode *clustering* terbaik di antara DBSCAN dan *K-means* yang dapat direkomendasikan untuk melakukan pengelompokan pada data faktor penyebab *stunting* di Indonesia dan data simulasi.

1.3 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah:

1. Penulis memperoleh hasil pengelompokan provinsi di Indonesia berdasarkan faktor penyebab *stunting* dan pengelompokan data simulasi menggunakan metode *Density-Based Spatial Clustering of Application with Noise* dan *K-means*.
2. Pembaca maupun penulis mendapat informasi mengenai metode *clustering* terbaik di antara DBSCAN dan *K-means* dalam melakukan pengelompokan.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data *mining* adalah proses penggalian data yang menghasilkan pengetahuan yang tidak disadari dan cenderung terabaikan (Nofriansyah & Gunadi, 2015). Data *mining* berasal dari beberapa gabungan ilmu pengetahuan seperti *machine learning*, statistika, dan sistem basis data. Istilah data *mining* atau *Knowledge Discovery Database* (KDD) seringkali digunakan untuk menjelaskan proses penggalian informasi yang tersembunyi dalam suatu basis data.

Menurut Nofriansyah & Gunaidi (2015), proses data *mining* atau *Knowledge Discovery Database* (KDD) memiliki beberapa fase yaitu:

1. Seleksi Data (*Selection*)

Pada tahap ini akan dilakukan seleksi data berdasarkan kriteria tertentu. Data yang paling sesuai dari sekian banyak data akan dipilih untuk diolah pada tahap berikutnya.

2. Pemilihan Data (*Cleaning*)

Proses ini meliputi membuang data yang mengandung duplikat, *missing value*, kolom yang tidak terpakai, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Menurut Zhang, *et al.* (2003), proses ini bertujuan untuk menghasilkan data yang berkualitas sehingga mengarah pada pola yang berkualitas.

3. Transformasi (*Transformation*)

Pada fase ini, data yang belum memiliki entitas yang jelas akan ditransformasi kedalam bentuk data yang valid sehingga data siap untuk digunakan pada proses data *mining*. Fase ini juga mencakup standarisasi data yang dapat menggunakan rumus sebagai berikut (Solimun, dkk., 2017):

$$Z_i = \frac{x_i - \bar{X}}{S_x} \quad (2.1.1)$$

dengan,

$$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}}$$

keterangan:

- x_i = data ke- i
- \bar{X} = rata-rata dari data x
- S_x = standar deviasi dari data
- n = jumlah data.

4. Data *Mining*

Fase ini memuat proses ekstraksi pola dari data yang ada dengan menggunakan algoritma atau metode-metode pencarian pengetahuan.

5. Interpretasi/Evaluasi (*Interpretation/Evaluation*)

Fase ini merupakan fase akhir dalam proses KDD. Pada fase ini pola akan diinterpretasikan kedalam bentuk pengetahuan yang mudah dimengerti dan dipahami oleh semua orang dan dapat digunakan untuk mendukung pengambilan keputusan.

Menurut Adinugroho & Yuita (2018), dalam pengoperasiannya data *mining* dapat dikelompokkan menjadi dua kategori yaitu metode deskriptif, dan metode prediktif. Metode deskriptif memiliki tujuan untuk menemukan pola, hubungan, atau anomali dalam data sehingga dapat mudah dipahami oleh manusia. Contoh dari metode deskriptif yaitu *clustering* dan *association rule*. Sedangkan, metode prediktif bertujuan untuk memperkiraan nilai dari suatu variabel dalam data

berdasarkan nilai variabel-variabel lainnya dari data tersebut. Contoh dari metode prediktif ini yaitu klasifikasi dan regresi.

2.2 Clustering

Clustering merupakan pengelompokan data yang memiliki kemiripan antara satu data dengan data lainnya ke dalam satu *cluster* sehingga data dalam satu *cluster* memiliki tingkat kemiripan (*similarity*) yang tinggi dan data antar *cluster* memiliki kemiripan yang rendah (Nishom, 2019). *Clustering* merupakan salah satu contoh dari metode deskriptif pada operasi dalam data *mining*. Menurut Purnomo, dkk. (2022), sebuah *cluster* dikatakan baik jika suatu *cluster* mempunyai anggota-anggota yang memiliki tingkat kesamaan karakteristik antar satu data dengan yang lainnya, tetapi sangat tidak memiliki kesamaan karakteristik dengan anggota-anggota pada *cluster* yang lain.

2.3 Jarak Euclidean

Ukuran jarak atau similiaritas digunakan untuk menerangkan keserupaan atau kedekatan antar data (Hasanah, 2022). Jarak Euclidean merupakan ukuran jarak yang sering digunakan dalam melakukan pengelompokan. Rumus jarak Euclidean adalah sebagai berikut (Akbar, *et al.*, 2023):

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.3.1)$$

keterangan (metode DBSCAN):

- $d(x_i, x_j)$ = jarak antara objek ke- i dan objek ke- j
- x_{ik} = objek ke- i pada variabel ke- k
- x_{jk} = objek ke- j pada variabel ke- k

keterangan (metode *K-Means*):

$d(x_i, x_j)$ = jarak antara objek ke- i dan *centroid*

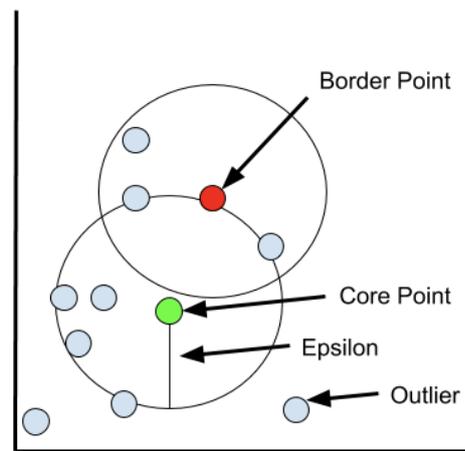
x_{ik} = objek ke- i pada variabel ke- k

x_{jk} = *centroid* pada variabel ke- k

2.4 *Density-Based Spatial Clustering of Application with Noise* (DBSCAN)

DBSCAN merupakan algoritma *clustering* berbasis kepadatan. Menurut Devi, dkk. (2015), DBSCAN adalah metode *clustering* yang membangun area pengelompokan berdasarkan kepadatan yang terkoneksi (*density-connected*) dimana pada setiap *cluster* akan memuat sejumlah minimum data dan mengidentifikasi data yang tidak masuk ke dalam *cluster* sebagai pencilan. Pada metode ini, jumlah *cluster* tidak ditentukan di awal. Algoritma metode ini akan melakukan pengelompokan berdasarkan tingkat kemiripan suatu data dengan data yang lainnya sehingga akan menghasilkan suatu *cluster* yang memiliki karakteristik yang sama di tiap *clusternya*.

Dalam prosesnya, metode ini memiliki dua parameter yaitu *min-points* dan epsilon yang harus ditentukan besarannya di awal proses. Menurut Prasetyo (2012), *min-points* memuat berapa banyaknya minimum data pada radius epsilon. Sedangkan epsilon merupakan radius atau jarak dari suatu poin ke titik terluar sehingga area tersebut akan membentuk satu *cluster*. Konsep kepadatan seperti ini menghasilkan tiga macam status dari setiap data, yaitu poin inti (*core*), poin tepi (*border*), dan poin pencilan. Poin inti merupakan suatu poin (p) dimana jumlah data lain dan p dalam radius ϵ berjumlah \geq *min-points* yang ditentukan. Poin tepi merupakan suatu poin (q) dimana dalam radius ϵ jumlah data lain dan $q <$ *min-points* yang ditentukan dan poin inti berada dalam radius ϵ tersebut. Sedangkan poin pencilan merupakan suatu poin (r) dimana jumlah data lain dan $r <$ *min-points* yang ditentukan dan tidak ada poin inti yang berada dalam radius.



Gambar 1. Parameter Metode DBSCAN
(Sumber: *Towards Data Science*)

Menurut Adha, dkk. (2021), berikut ini merupakan tahap-tahap untuk melakukan *clustering* dengan menggunakan metode DBSCAN:

1. Menentukan nilai *min-points* dan epsilon
2. Menentukan titik awal (p) secara acak
3. Hitung semua jarak titik yang *density reachable* terhadap p dengan menggunakan rumus jarak Euclidean
4. *Cluster* akan terbentuk jika titik yang memenuhi epsilon berjumlah \geq *min-points*, dan titik p akan diidentifikasi sebagai poin inti.
5. Lakukan pengulangan langkah 2 – 4 hingga dilakukan proses pada semua titik. Proses dilanjutkan ke titik yang lain apabila p merupakan poin tepi dan tidak ada titik yang *density reachable* terhadap p .

2.5 K-Means

K-Means merupakan salah satu algoritma *clustering* yang berbasis partisi yang berusaha mempartisi data yang ada ke dalam *cluster-cluster* (Mubarak & Gusmi, 2023). Setiap *cluster* pada metode *K-Means* memiliki karakteristik yang berbeda dikarenakan objek pada satu *cluster* tidak mirip dengan objek-objek milik *cluster*

lain (Nirmal, 2019). Algoritma ini menghitung jarak antara data dengan setiap pusat *cluster* dengan menggunakan persamaan jarak Euclidean. Metode ini membutuhkan jumlah *cluster* yang ditentukan terlebih dahulu sebagai k , dan kemudian membagi n objek ke dalam k *cluster*.

Algoritma pengelompokan ini membutuhkan jumlah *cluster* k yang sesuai, karena *centroid* awal dapat berubah-ubah, yang berdampak pada hasil pengelompokan data yang tidak konsisten (Setiawan, *et al.*, 2023). *Centeroid* dibentuk oleh pemilihan acak dari objek pada *cluster* (Arora, *et al.*, 2016). Namun *centeroid* pada proses selanjutnya akan dihitung kembali dengan menggunakan rumus sebagai berikut (Ediyanto & Satyahadewi, 2021):

$$V = \frac{1}{n} \sum_{i=1}^n x_i ; i = 1, 2, \dots, n \quad (2.5.1)$$

keterangan:

- V = *centeroid* pada *cluster*
- n = jumlah objek yang menjadi anggota *cluster*
- x_i = nilai objek ke- i pada *cluster* tersebut

Menurut Purnomo, dkk. (2022), berikut ini merupakan tahap-tahap untuk melakukan *clustering* dengan menggunakan metode *K-Means*:

1. Menentukan banyaknya *cluster* dengan menggunakan metode *elbow*.
2. Menentukan data yang akan dijadikan pusat *cluster* (*centeroid*).
3. Menghitung jarak antar objek dan masing-masing *centeroid* dengan menggunakan rumus jarak Euclidean.
4. Mengelompokkan objek berdasarkan jarak terdekat dengan *centeroid*.
5. Menentukan *centeroid* baru dengan menggunakan persamaan (2.5.1)
6. Mengulangi langkah 3 dan 4 sampai tidak ada lagi data yang berpindah *cluster*.

2.6 Silhouette Coefficient

Silhouette coefficient merupakan metode yang digunakan dalam menggambarkan seberapa tepat suatu objek data yang ditempatkan pada sebuah *cluster* (Paembonan & Hisma, 2021). Perhitungan *Silhouette coefficient* bertujuan untuk mengetahui kualitas dari pengelompokan atau pengukuran ketepatan pengelompokan. Untuk menghitung nilai *Silhouette coefficient* dapat menggunakan persamaan berikut (Akbar, *et al.*, 2023):

$$SIC = \frac{1}{n} \sum_{i=1}^n \left(\frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \right) \quad (2.6.1)$$

dimana,

$$a(i) = \frac{1}{n_A - 1} \sum_{i,j \in A, i \neq j} d(i, j)$$

$$b(i) = \min \left(\frac{1}{n_C} \sum_{p \in C} d(i, p) \right)$$

keterangan:

- n_A = jumlah data pada *cluster A*
- n_C = jumlah data pada *cluster C* dimana $C \neq A$
- $d(i, j)$ = jarak objek ke- i dengan objek lainnya pada *cluster* yang sama
- $d(i, p)$ = jarak objek ke- i dengan objek lainnya pada *cluster* yang berbeda
- $a(i)$ = jarak rata-rata sampel i ke sampel lain dalam *cluster*
- $b(i)$ = jarak minimum sampel i ke *cluster* lain.

Kriteria pengelompokan berdasarkan *Silhouette coefficient* dapat dilihat pada Tabel 1 sebagai berikut:

Tabel 1. Pengelompokan berdasarkan *Silhouette coefficient* (Kaufman & Rousseeuw, 2009).

Nilai SIC	Kriteria
0.71-1.00	Struktur Kuat
0.51-0.70	Struktur Baik
0.26-0.50	Struktur Lemah
≤ 0.25	Struktur Buruk

2.7 Calinski-Harabasz (CH)

Calinski-Harabasz merupakan salah satu metode validitas *cluster* yang bertujuan untuk mengevaluasi kualitas dari klasterisasi. Metode *Calinski-Harabasz* didasarkan pada hubungan antar *Between Cluster Sum of Square* (BCSS) dan *Within-Cluster Sum of Square* (WCSS) dengan rumus (Wang & Xu, 2019):

$$CH = \frac{(BCSS)(N - k)}{(WCSS)(k - 1)} \quad (2.7.1)$$

dimana,

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} d(x_i, \bar{x}_k)^2$$

$$BCSS = \sum_{k=1}^K n_k d(\bar{x}_k, \bar{x})^2$$

keterangan:

- k = banyaknya *cluster*
- N = banyaknya titik data
- n_k = jumlah titik data dalam *cluster* ke- k
- C_k = *cluster* ke- k

$$d(x_j, \bar{x}_k)^2 = \text{jarak antara titik data } (x_i) \text{ dan pusat data dari cluster } k (\bar{x}_k)$$

$$d(\bar{x}_k, \bar{x})^2 = \text{jarak antara pusat data setiap cluster } (\bar{x}_k) \text{ dan pusat data dari seluruh dataset } (\bar{x})$$

2.8 Metode *Elbow*

Menurut Madhulatha (2012), metode *elbow* merupakan metode yang bekerja dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik dalam mendapatkan informasi yang akan digunakan untuk menentukan jumlah *cluster* terbaik pada suatu data. Presentase dari setiap jumlah *cluster* tersebut dapat dibuat dalam bentuk grafik dalam memudahkan pencarian informasi. Nilai *cluster* terbaik ditunjukkan melalui grafik dimana jumlah *cluster* satu dengan *cluster* selanjutnya akan memberikan sudut atau dengan kata lain *cluster* satu dengan *cluster* selanjutnya akan mengalami penurunan nilai yang paling besar.

Untuk memperoleh perbandingannya dapat menggunakan metrik *Within Cluster Sum of Squares* (WCSS). WCSS umumnya digunakan pada metode *elbow* dalam menentukan jumlah *cluster* yang optimal dalam pengelompokan *K-means*. WCSS adalah jumlah kuadrat jarak antara setiap titik dan pusat data dalam sebuah *cluster* (Harmain, dkk., 2021). Semakin besar jumlah *cluster*, maka nilai WCSS akan semakin berkurang. WCSS dapat dihitung dengan rumus sebagai berikut (Irwanto, dkk., 2012):

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} d(x_i, \bar{x}_k)^2 \quad (2.8.1)$$

keterangan:

- k = banyaknya *cluster*.
- C_k = *cluster* ke- k
- \bar{x}_k = *centeroid* dari *cluster* C_k

x_i = data ke- i pada *cluster* C_k

$d(x_i, \bar{x}_k)$ = jarak Euclidean antara data x_i dan *centroid cluster* \bar{x}_k

2.9 Stunting

Mengutip Imani (2020), menurut dr. Fatimah Hidayati, Sp.A *stunting* adalah kondisi dimana anak lebih pendek dibandingkan dengan anak lain seusianya, atau dengan kata lain tinggi badan anak berada di bawah standar. Hal ini dapat disebabkan oleh asupan gizi yang kurang terutama pada 1.000 Hari Pertama Kehidupan (HPK). Kondisi kekurangan asupan gizi ini berlangsung sejak bayi masih dalam kandungan dan baru akan nampak setelah bayi berusia 2 tahun.

Banyak faktor yang menyebabkan *stunting* terjadi baik faktor langsung maupun faktor tidak langsung. Menurut Nuryuliyani (2023), faktor-faktor penyebab *stunting* antara lain yaitu asupan gizi dan status kesehatan yang meliputi ketahanan pangan (ketersediaan, keterjangkauan dan akses pangan bergizi), lingkungan sosial (norma, makanan bayi dan anak, *hygiene*, pendidikan, dan tempat kerja), lingkungan kesehatan (akses, pelayanan preventif dan kuratif), dan lingkungan pemukiman (air, sanitasi, kondisi bangunan).

3. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2023/2024 dan bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini terdiri dari 2 data yaitu data real dan data simulasi. Data real yang digunakan berupa faktor-faktor yang mempengaruhi *stunting* di seluruh provinsi Indonesia pada tahun 2022 dengan variabel-variabel yang digunakan yaitu: pemberian ASI eksklusif (X_1), rumah tangga dengan akses sanitasi layak (X_2), proses kelahiran dengan bantuan tenaga kesehatan terlatih (X_3), dan proses kelahiran di fasilitas kesehatan (X_4). Data tersebut merupakan data sekunder yang diperoleh melalui situs resmi Badan Pusat Statistik (BPS) pada link berikut: <https://www.bps.go.id/>. Sedangkan untuk data simulasi, data dibangkitkan dari distribusi normal dengan 4 variabel X_1, X_2, X_3, X_4 . Pada *cluster 1*, $X_j \sim N(1,2)$ dengan $j = 1,2,3,4$ dan jumlah data sebanyak 50 data. Pada *cluster 2* $X_j \sim N(7,2)$ dengan $j = 1,2,3,4$ dan jumlah data sebanyak 50 data. Data dari kedua *cluster* digabungkan sehingga didapatkan dataset sebanyak 100 yang akan dilakukan pengelompokkan.

3.3 Metode Penelitian

Metode yang digunakan pada penelitian ini adalah *Density-Based Spatial Clustering of Application with Noise* dan *K-means*. Tahapan dari melakukan perbandingan antara kedua metode tersebut sebagai berikut:

1. Analisis deskriptif terhadap data real dan data simulasi.
2. Melakukan transformasi data yaitu melakukan standarisasi data dengan menggunakan persamaan (2.1.1) pada data real dan data simulasi.
3. Melakukan *clustering* dengan menerapkan algoritma DBSCAN sebagai berikut:
 1. Menentukan nilai *min-points* dan epsilon yaitu nilai *min-points* 2, 3, 4, dan 5 dan nilai epsilon yaitu: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, dan 15 untuk data faktor penyebab *stunting*, dan nilai *min-points* 2, 3, 4, dan 5 juga nilai epsilon yaitu 0,5; 1; 1,5; dan 2 pada data simulasi
 2. Menentukan titik awal (p) secara acak
 3. Hitung semua jarak titik yang *density reachable* terhadap p berdasarkan nilai epsilon yang digunakan dengan menggunakan rumus jarak Euclidean pada persamaan (2.3.1)
 4. *Cluster* akan terbentuk jika titik yang memenuhi epsilon dan berjumlah \geq *min-points*, dan titik p akan diidentifikasi sebagai poin inti.
 5. Lakukan pengulangan langkah 2 – 4 hingga dilakukan proses pada semua titik. Proses dilanjutkan ke titik yang lain apabila p merupakan poin tepi dan tidak ada titik yang *density reachable* terhadap p .
 6. Lakukan pengulangan langkah 1-5 pada seluruh nilai *min-points* dan epsilon yang ditentukan.
4. Melakukan *clustering* dengan menerapkan algoritma *K-means* sebagai berikut:
 1. Menentukan banyaknya *cluster* dengan beberapa percobaan nilai k yaitu 2, 3, 4, dan 5.
 2. Menentukan data yang akan dijadikan pusat *cluster* (*centroid*).

3. Menghitung jarak antar objek dan masing-masing *centroid* dengan menggunakan rumus jarak Euclidean pada persamaan (2.3.1)
 4. Mengelompokkan objek berdasarkan jarak terdekat dengan *centroid*.
 5. Menentukan *centroid* baru dengan menggunakan persamaan (2.5.1)
 6. Mengulangi langkah 3 dan 4 sampai tidak ada lagi data yang berpindah *cluster*.
 7. Lakukan pengulangan langkah 1-6 pada semua nilai k .
5. Evaluasi hasil *clustering* metode DBSCAN dan *K-Means* dengan melakukan uji validitas yaitu menghitung nilai *Silhouette coefficient* (persamaan 2.6.1) dan nilai *Calinski-Harabasz* (persamaan 2.7.1) pada hasil *cluster* kedua metode tersebut.
 6. Membandingkan hasil uji validitas metode DBSCAN dan *K-Means* berdasarkan nilai *Silhouette coefficient* dan *Callinski-Harabasz* yang dihasilkan.

5. KESIMPULAN

Berdasarkan hasil penelitian mengenai klusterisasi metode DBSCAN dan *K-means* dalam pengelompokan provinsi berdasarkan faktor penyebab *stunting* di Indonesia diperoleh kesimpulan bahwa metode *K-means* menghasilkan pengklasteran yang lebih baik dari metode DBSCAN berdasarkan nilai *Silhouette coefficient* yaitu sebesar 0,53 dan nilai *Callinski-Harabasz* sebesar 26,09. Selain itu, baik metode DBSCAN maupun metode *K-means* menghasilkan kluster berjumlah dua. Namun dalam metode DBSCAN terdapat dua data yang diidentifikasi sebagai pencilan sehingga kedua data tersebut dipisahkan dalam kluster tersendiri. Dalam metode DBSCAN dan *K-means cluster 1* memiliki karakteristik faktor-faktor penyebab *stunting* dengan presentase setiap variabelnya lebih tinggi, sedangkan *cluster 2* memiliki faktor-faktor penyebab *stunting* yang relatif lebih rendah.

Namun, hasil yang berbeda diperoleh ketika dilakukan *clustering* menggunakan data simulasi dimana data tersebut berdistribusi normal dan tidak mengandung pencilan pada data. Hasil penelitian menunjukkan bahwa dengan menggunakan data simulasi, metode DBSCAN menghasilkan pengklasteran yang lebih baik dari metode *K-means* berdasarkan nilai *Silhouette coefficient* yaitu sebesar 0,65 dan nilai *Callinski-Harabasz* sebesar 328,03. Selain itu, baik metode DBSCAN maupun metode *K-means* menghasilkan kluster berjumlah dua. Dalam metode DBSCAN dan *K-means cluster 1* memiliki karakteristik data yang nilainya lebih rendah dibandingkan dengan *cluster 2*.

Penelitian ini menjelaskan bahwa performa dari metode DBSCAN dan *K-Means* dipengaruhi oleh data yang digunakan. Pada penelitian ini, metode DBSCAN memiliki performa yang baik untuk data yang tidak mengandung pencilan, dan metode *K-means* memiliki performa yang lebih baik untuk data yang mengandung pencilan.

DAFTAR PUSTAKA

- Adha, R., Nana, N., Umami, S., & Mustakim, M. 2021. Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*. **18**(2): 206-211.
- Adinugroho, S., & Yuita, A. S. 2018. *Implementasi Data Mining Menggunakan Weka*. UB Press, Malang.
- Akbar, T., Tinungki, G. M., & Siswanto, S. 2023. Performance Comparison of K-Medoids and Density Based Spatial Clustering of Application with Noise Using Silhouette Coefficient Test. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*. **17**(3): 1605-1616.
- Arora, P., Deepali., & Shipra, V. 2016. Analysis of K-Means and K-medoids Algorithm for Big Data, hlm. 507-512. International Conference on Information Security & Privacy, Nangpur, India.
- Devi, A. S., Putra, I. K. G. D., & Sukarsa, I. M. 2015. Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan. *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*. **6**(3): 185.
- Harmain, A., Paiman, P., Henri, K., Kusriani, K., & Dina, M. 2021. Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas. *TEKNIMEDIA: Teknologi Informasi dan Multimedia*. **2**(2): 83-89.
- Hasanah, I. N. 2022. Analisis Cluster Berdasarkan Dampak Ekonomi Di Indonesia Akibat Pandemi Covid-19. *MATHunesa: Jurnal Ilmiah Matematika*. **10**(2): 239-248.

- Imani, N. 2020. *Stunting Pada Anak: Kenali dan Cega Sejak Dini*. Hijaz Pustaka Mandiri, Yogyakarta.
- Irwanto, I., Purwananto, Y., & Soelaiman, R. 2012. Optimasi Kinerja Algoritma Klusterisasi K-Means untuk Kuantisasi Warna Citra. *Jurnal Teknik ITS*. **1**(1): 197-202.
- Jollyta, D., Wiliam, R., & Muhammad, Z. 2020. *Konsep Data Mining dan Penerapan*. Deepublish, Yogyakarta.
- Kaufman, L. & Rousseeuw, P. J. 2009. *Finding Groups in Data: an Introduction to Cluster Analysis*. John Wiley & Sons, New Jersey.
- Kristianto, A. 2021. Analisa Performa K-Means dan DBSCAN dalam Clustering Minat Penggunaan Transportasi Umum. *Elkom: Jurnal Elektronika dan Komputer*. **14**(2): 368-372.
- Li, J., Anna, Z., Wei, G., Nairwita, B., Yanji, Z., & Qianfeng, W. 2023. Urban Flood Risk Assessment Based on DBSCAN and K-Means Clustering Algorithm. *Geomatics, Natural Hazards and Risk*. **14**(1): 1-28.
- Madhulatha, T. S. 2012. An Overview on Clustering Methods. *IOSR Journal of Engineering*. **2**(4): 719-725.
- Matdoan, M. Y., Matdoan, U, A., & Far-Far, M. S. 2022. Algoritma K-Means Untuk Klasifikasi Provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting. *PANRITA: Journal of Science, Technology, and Arts*. **1**(2): 41-46.
- Mubarak, H. & Gusmi, K. 2023. Analisis Cluster Dalam Pengelompokan Kabupaten/Kota Di Provinsi Jambi Berdasarkan Penyakit Menular Menggunakan Metode K-Means. *Jurnal Statistika dan Komputasi*. **2**(1): 20-30.
- Muflikhah, L., Dian, E. R. & Rekyan, R. M. P. 2018. *Data Mining*. UB Press, Malang. 182 hlm.

- Ediyanto, M. N. M., & Satyahadewi, N. 2013. Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*. **2**(02): 133-136.
- Nirmal, S. 2019. Comparative Study Between K-Means and K-medoids Clustering Algorithms. *J. Classif.* **6**(3): 839-844.
- Nishom, M. 2019. Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika*. **4**(1): 20-24.
- Nofriansyah, D. & Gunadi, W. N. 2015. *Algoritma Data Mining dan Pengujian*. Deepublish, Yogyakarta. 118 hlm.
- Nuryuliyani, E. 2023. Kementerian Kesehatan Republik Indonesia. Mengenal Lebih Jauh Tentang Stunting. https://yankes.kemkes.go.id/view_artikel/2657/mengenal-lebih-jauh-tentangstunting#:~:text=Penyebab%20stunting%20antara%20lain%20yaitu,pelayanan%20preventif%20dan%20kuratif%2C%20dan. Diakses pada 26 November 2023.
- Onis, D. M. & Francesco, B. 2016. Childhood Stunting: A Global Perspective. *Maternal & Child Nutrition*. **12**: 12-26.
- Paembonan, S. & Hisma, A. 2021. Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*. **6**(2): 48-54.
- Paramashanti, B. A. 2019. *Gizi Bagi Ibu dan Anak*. Pustaka Baru Press, Yogyakarta.
- Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. CV Andi Offset, Yogyakarta.
- Purnomo., Eddy, S., Wargijono, U., Okta, P., Rachmat, F., Sulistianingsih, A. S., Ratna, F., Annisa, C., & Najih, G. R. 2022. *Analisis Data Multivariat*. Omera Pustaka, Jawa Tengah.

- Setiawan, K. E., Afdhal, K., Andry, C., & Derwin, S. 2023. Clustering Models for Hospitals in Jakarta using Fuzzy C-Means and K-Means. *Procedia Computer Science*. **216**(3): 356-363.
- Solimun, MS., Adji, A. R. F., & Nurjannah. 2017. *Metode Statistika Multivariat: Pemodelan Persamaan Struktural (SEM) Pendekatan WarpPLS*. UB Press, Malang.
- Wang, X., & Xu, Y. 2019. An Improved Index for Clustering Validation Based on Silhouette Index and Calinski-Harabasz Index. *IOP Conference Series: Materials Science and Engineering*. **569**(5): 1-6.
- Zhang, S., Chengqi, Z., & Qiang, Y. 2003. Data Preparation for Data Mining. *Applied artificial intelligence*. **17**(5-6): 375-381.