

SKRIPSI

**CLUSTERING STATUS GIZI BAYI MENGGUNAKAN
DENSITY BASED SPATIAL CLUSTERING OF
APPLICATIONS WITH NOISE (DBSCAN)
(STUDI KASUS PUSKESMAS SUDIANG KOTA MAKASSAR)**

Disusun dan diajukan oleh:

**MOCH. WAHYU FAISAL
D121 17 1516**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024**

LEMBAR PENGESAHAN SKRIPSI

**CLUSTERING STATUS GIZI BAYI MENGGUNAKAN DENSITY
BASED SPATIAL CLUSTERING OF APPLICATIONS WITH
NOISE (DBSCAN)
(STUDI KASUS PUSKESMAS SUDIANG KOTA MAKASSAR)**

Disusun dan diajukan oleh

**Moch. Wahyu Faisal
D121 17 1516**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian
Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada tanggal 5 April 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,

Dr. Ir. Ingrid Nurtanio, M.T.
NIP 19610813 198811 2 001

Prof. Dr. Eng. Intan Sari Areni, S.T.,
M.T.
NIP 19750203 200012 2 002

Ketua Program Studi,



Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.
NIP 19750716 200212 1 004



PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama : Moch. Wahyu Faisal

NIM : D121 17 1516

Program Studi : Teknik Informatika

Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Clustering Status Gizi Bayi Menggunakan Density Based Spatial Clustering of Applications With Noise (DBSCAN) (Studi Kasus Puskesmas Sudiang Kota Makassar)

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gōwa, 5 April 2024



ng Menyatakan

Moch. Wahyu Faisal



ABSTRAK

MOCH. WAHYU FAISAL. *Clustering Status Gizi Bayi Menggunakan Density Based Spatial Clustering of Applications With Noise (DBSCAN) (Studi Kasus Puskesmas Sudiang Kota Makassar)* (dibimbing oleh Ingrid Nurtanio dan Intan Sari Areni)

Status gizi bayi merupakan indikator penting dalam menilai kesehatan dan kesejahteraan anak. Status gizi didefinisikan sebagai keadaan tubuh sebagai akibat konsumsi makanan dan penggunaan zat-zat gizi. Salah satu metode yang dapat digunakan untuk mengetahui status gizi bayi dengan menggunakan metode *clustering*. *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) adalah sebuah algoritma *clustering* yang mengelompokkan objek data berdasarkan kepadatannya, yang mana sebuah kluster dianggap sebagai kelompok yang memiliki objek yang padat, dan dipisahkan oleh kelompok yang memiliki kepadatan objek yang rendah atau *noise*.

Tujuan dari penelitian ini adalah untuk membangun sistem pengelompokkan status gizi bayi dengan metode DBSCAN dan memberikan informasi bagaimana kerja sistem tersebut dalam mengelompokkan status gizi bayi dengan metode DBSCAN.

Pada penelitian ini, penerapan metode DBSCAN dilakukan melalui tahapan-tahapan yang meliputi *input data*, *preprocessing data*, implementasi metode, menentukan nilai Eps dan MinPts, dan melakukan analisis hasil kluster yang terbentuk.

Hasil penelitian dengan Eps = 0,3 dan MinPts = 6 menghasilkan kluster sebanyak 8 dengan nilai *Silhouette Coefficient* = 0,7319 dan SSE = 6,1343. Sebanyak 18 bayi berada pada kluster 0, sebanyak 19 bayi berada pada kluster 1, sebanyak 10 bayi berada pada kluster 2, sebanyak 19 berada pada kluster 3, sebanyak 18 bayi berada pada kluster 4, sebanyak 29 bayi berada pada kluster 5, sebanyak 16 bayi berada pada kluster 6, dan sebanyak 15 bayi berada pada kluster 7. Dengan diketahuinya hasil kluster ini dapat dijadikan sebagai informasi identifikasi awal dan pencegahan masalah gizi bayi sehingga dapat meningkatkan status gizi bayi.

Kata Kunci: *Clustering*, DBSCAN, Status Gizi Bayi



ABSTRACT

MOCH. WAHYU FAISAL. *Clustering Infant Nutrition Status Using Density Based Spatial Clustering of Applications With Noise (DBSCAN) (Case Study of Puskesmas Sudiang, Makassar)* (supervised by Ingrid Nurtanio and Intan Sari Areni)

The nutritional status of infants is an important indicator in assessing the health and well-being of children. Nutritional status is defined as the state of the body as a result of food consumption and the use of nutrients. One method that can be used to determine the nutritional status of infants using clustering methods. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering algorithm that groups data objects based on their density, where a cluster is considered a group that has dense objects, and is separated by a group that has a low density of objects or noise.

The purpose of this research is to build an infant nutritional status clustering system with the DBSCAN method and provide information on how the system works in classifying the nutritional status of infants using the DBSCAN method.

In this study, the application of the DBSCAN method is carried out through stages including data input, data preprocessing, method implementation, determining the value of Eps and MinPts, and analyzing the results of the clusters formed.

The results with $Eps = 0.3$ and $MinPts = 6$ resulted in 8 clusters with Silhouette Coefficient = 0.7319 and $SSE = 6.1343$. A total of 18 infants are in cluster 0, 19 infants are in cluster 1, 10 infants are in cluster 2, 19 are in cluster 3, 18 infants are in cluster 4, 29 infants are in cluster 5, 16 infants are in cluster 6, and 15 infants are in cluster 7. Knowing the results of this cluster can be used as information for early identification and prevention of infant nutrition problems so as to improve the nutritional status of infants.

Keywords: Clustering, DBSCAN, Infant Nutritional Status



DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	i
PERNYATAAN KEASLIAN.....	ii
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vi
DAFTAR TABEL.....	vii
DAFTAR SINGKATAN DAN ARTI SIMBOL	viii
DAFTAR LAMPIRAN.....	ix
KATA PENGANTAR	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	3
1.5 Ruang Lingkup.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Status Gizi	4
2.2 <i>Data Mining</i>	5
2.3 <i>Clustering</i>	7
2.4 <i>Principal Component Analysis (PCA)</i>	8
2.5 <i>Euclidian Distance</i>	11
2.6 <i>Density Based Spatial Clustering of Applications with Noise (DBSCAN)</i> 12	
2.7 <i>Silhouette Coefficient</i>	17
2.8 <i>Sum of Squared Error (SSE)</i>	18
2.9 Pemetaan	18
BAB 3 METODE PENELITIAN/PERANCANGAN	20
3.1 Tahapan Penelitian.....	20
3.2 Waktu dan Lokasi Penelitian	21
3.3 Instrumen Penelitian.....	21
3.4 Teknik Pengambilan Data.....	21
3.5 Perancangan Sistem	23
3.6 Visualisasi Pemetaan.....	39
BAB 4 HASIL DAN PEMBAHASAN.....	40
4.1 Hasil Penelitian	40
4.2 Pembahasan.....	59
BAB 5 KESIMPULAN DAN SARAN.....	68
5.1 Kesimpulan	68
5.2 Saran.....	69
DAFTAR PUSTAKA	70



DAFTAR GAMBAR

Gambar 1 Tahap <i>Data Mining</i> (Han dkk., 2011).....	6
Gambar 2 Karakteristik sebuah cluster (Bhatia, 2019).....	7
Gambar 3 Ilustrasi proses PCA (Tharwat, 2016).....	9
Gambar 4 Ilustrasi pembentukan ruang PCA (Tharwat, 2016).....	11
Gambar 5 Konsep dari <i>density-reachability</i> (Bhatia, 2019).....	13
Gambar 6 <i>Pseudocode</i> algoritma DBSCAN (Han dkk., 2011).....	14
Gambar 7 Ilustrasi <i>ordered k-dist-graph</i> (Ester dkk., 1996).....	15
Gambar 8 Tahapan Penelitian.....	20
Gambar 9 Data bayi Puskesmas Sudiang.....	22
Gambar 10 Data tambahan variabel.....	22
Gambar 11 Rancangan sistem.....	24
Gambar 12 Sampel data riwayat pengukuran bayi.....	25
Gambar 13 Proses data <i>cleaning</i>	25
Gambar 14 Sampel data kategorik <i>transformation</i>	26
Gambar 15 Sampel data numerik <i>transformation</i>	27
Gambar 16 Sampel <i>dataframe mean-centering data (171x22)</i>	27
Gambar 17 Sampel <i>dataframe covariance matrix (22x22)</i>	28
Gambar 18 <i>Eigenvector</i> dari PCA (22x3).....	30
Gambar 19 Sampel hasil proses reduksi dimensi data.....	31
Gambar 20 Sampel data dengan PCA <i>ordered k-dist graph</i>	32
Gambar 21 <i>Ordered k-dist graph</i> data dengan PCA.....	33
Gambar 22 Sampel data tanpa PCA <i>ordered k-dist graph</i>	33
Gambar 23 <i>Ordered k-dist graph</i> data dengan PCA.....	34
Gambar 24 Flowchart algoritma DBSCAN.....	35
Gambar 25 Sampel data klaster 6.....	37
Gambar 26 Sampel perhitungan nilai a dan b pada <i>Silhouette Score</i>	38
Gambar 27 Sampel perhitungan <i>Silhouette Score</i>	38
Gambar 28 Visualisasi 2D dan 3D <i>dataset</i>	40
Gambar 29 Sampel hasil klasterisasi DBSCAN.....	41
Gambar 30 Visualisasi hasil klasterisasi.....	42
Gambar 31 Visualisasi pemetaan hasil <i>clustering</i>	66
Gambar 32 Pemetaan posyandu Bahagia.....	66



DAFTAR TABEL

Tabel 1 Interpretasi <i>Silhouette Coefficient</i> (Kaufman & Rousseeuw, 1990).....	17
Tabel 2 Uji coba Eps dan MinPts dengan PCA	40
Tabel 3 Uji coba Eps dan MinPts tanpa PCA	42
Tabel 4 Hasil 8 kluster.....	50
Tabel 5 Data Non Kluster.....	53
Tabel 6 Klaster Gizi Baik.....	60
Tabel 7 Klaster Risiko Gizi Lebih	63
Tabel 8 Klaster Gizi Lebih.....	64



DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
DBSCAN	<i>Density Based Spatial Clustering of Application with Noise</i>
Eps	<i>Epsilon</i>
MinPts	<i>Minimum Points</i>
PCA	<i>Principal Component Analysis</i>
SC	<i>Silhouette Coefficient</i>
SSE	<i>Sum of Squared Error</i>
CSV	<i>Comma Separated Values</i>



DAFTAR LAMPIRAN

Lampiran 1 Lembar Perbaikan Skripsi	73
Lampiran 2 Surat Izin Pengambilan Data Kampus	74
Lampiran 3 Surat Izin Pengambilan Data Dinas Kesehatan Kota Makassar	75
Lampiran 4 Contoh Dataset Bayi Puskesmas Sudiang Kota Makassar 2022	76
Lampiran 5 Tabel Hasil Klasterisasi Status Gizi Bayi Puskesmas Sudiang	78
Lampiran 6 Pemetaan Posyandu Kelurahan Sudiang	80
Lampiran 7 Pemetaan Posyandu Kelurahan Pai	81
Lampiran 8 Pemetaan Posyandu Kelurahan Laikang	82



KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Segala puji dan syukur kami panjatkan ke hadirat Allah SWT. Tuhan Yang Maha Esa yang dengan limpahan rahmat dan hidayah-Nya sehingga tugas akhir dengan judul “*Clustering Status Gizi Bayi Menggunakan Density Based Spatial Clustering of Application with Noise (DBSCAN) (Studi Kasus Puskesmas Sudiang Kota Makassar)*” ini dapat diselesaikan sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 pada Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Dalam penyusunan penelitian ini disajikan hasil penelitian terkait judul yang telah diangkat dan telah melalui proses pencarian dari berbagai sumber baik jurnal penelitian, prosiding pada seminar-seminar nasional/internasional, buku maupun dari situs-situs di internet.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai dengan masa penyusunan Tugas Akhir, sangatlah sulit untuk menyelesaikan Tugas Akhir ini. Oleh karena itu, penulis berterima kasih kepada:

- 1) Kedua Orang tua dan saudara-saudara penulis, yang selalu memberikan dukungan, doa dan semangat;
- 2) Ibu Dr. Ir. Ingrid Nurtanio, M.T. selaku pembimbing I dan Ibu Prof. Dr. Eng. Intan Sari Areni, S.T., M.T. selaku pembimbing II, yang senantiasa menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa dalam mengarahkan penulis untuk menyelesaikan tugas akhir;
- 3) Bapak Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN. Eng., selaku Ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas bimbingannya selama masa perkuliahan penulis;
- 4) Para Sahabat, teman-teman yang telah memberikan begitu banyak bantuan, semangat, inspirasi, ilmu dan hiburan selama penelitian, pengambilan data dan diskusi *progress* penyusunan Tugas Akhir;



- 5) Segenap Staf Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah membantu penulis;
- 6) Seluruh teman-teman angkatan 2017 Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin;
- 7) Orang-orang berpengaruh lainnya yang tanpa sadar telah menjadi inspirasi penulis.

Akhir kata, penulis berharap semoga Allah SWT. berkenan membalas segala kebaikan dari semua pihak yang telah banyak membantu. Semoga Tugas Akhir ini dapat memberikan manfaat bagi pengembangan ilmu. Aamiin.

Wassalam

Gowa, 5 April 2024

Penulis



BAB I PENDAHULUAN

1.1 Latar Belakang

Status gizi bayi merupakan indikator penting dalam menilai kesehatan dan kesejahteraan anak. Status gizi didefinisikan sebagai keadaan tubuh sebagai akibat konsumsi makanan dan penggunaan zat-zat gizi (Iskandar, 2022). Definisi ini menekankan pentingnya evaluasi kondisi gizi untuk memastikan pertumbuhan dan perkembangan yang optimal pada bayi. Status gizi yang baik dibutuhkan untuk mencegah berbagai masalah kesehatan dan perkembangan yang mungkin timbul akibat kekurangan atau kelebihan nutrisi.

Data mining adalah proses menemukan pola dan pengetahuan yang menarik dari sejumlah data yang besar (Han dkk., 2011). Pendekatan ini memungkinkan penemuan pengetahuan yang tidak diketahui sebelumnya dan pengenalan pola yang dapat membantu dalam pengambilan keputusan. Pemanfaatan *data mining* juga dapat membantu dalam proses pengelompokan data berdasarkan kemiripan karakteristik sebuah data. *Clustering* adalah salah satu teknik dalam *data mining* yang bertujuan untuk mengelompokkan data-data (objek) ke dalam beberapa kluster atau kelompok sehingga objek yang serupa disatukan ke dalam kluster yang sama, dan berbeda harus menjadi bagian dari kluster yang berbeda (Benabdellah dkk., 2019).

Salah satu algoritma dalam pengelompokan atau *clustering* yang sering digunakan adalah algoritma DBSCAN. *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) adalah sebuah algoritma *clustering* yang mengelompokkan objek data berdasarkan kepadatannya, yang mana sebuah kluster dianggap sebagai kelompok yang memiliki objek yang padat, dan dipisahkan oleh kelompok yang memiliki kepadatan objek yang rendah atau *noise* (Karami & Johansson, 2015).

Penelitian terkait penggunaan *clustering* dalam pengelompokan status gizi



ah dilakukan dengan algoritma K-Means oleh Winiarti dkk., pada tahun 2018. Hasil penelitian yang didapatkan dengan menggunakan *Cross-Validation* menunjukkan validasi 90% bahwa sistem dapat menentukan status gizi balita dengan

memproduksi 5 kluster. Penelitian terkait penggunaan DBSCAN dalam pengelompokan data juga telah dilakukan pada lahan pertanian Padi di Kabupaten Karawang. Hasil eksperimen pada penelitian ini menunjukkan silhouette score tertinggi sebanyak 0,74 dengan dua kluster (Nurina Sari dkk., 2019).

Pemetaan adalah proses pengumpulan data untuk dijadikan sebagai langkah awal dalam pembuatan peta, dengan menggambarkan penyebaran kondisi alamiah tertentu secara meruang, memindahkan keadaan sesungguhnya kedalam peta dasar, yang dinyatakan dengan penggunaan skala peta. (Mudhari, 2018).

Penelitian ini akan mengimplementasikan *Density Based Spatial Clustering of Application with Noise* (DBSCAN) dengan tujuan untuk mengelompokkan status gizi bayi menjadi beberapa kelompok berdasarkan kemiripan karakteristik yang ditinjau dari indikator kesehatan yang relevan. Hasil akhir penelitian ini akan diwujudkan ke dalam bentuk visualisasi informasi yang diharapkan dapat memudahkan pemahaman tentang distribusi status gizi bayi dan juga memberikan wawasan yang dapat dijadikan dasar dalam pengambilan keputusan intervensi gizi.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini diuraikan sebagai berikut:

- a. Bagaimana membangun sistem pengelompokkan status gizi bayi wilayah kerja Puskesmas Sudiang dengan *Density Based Spatial Clustering of Application with Noise* (DBSCAN)?
- b. Bagaimana unjuk kerja sistem pengelompokkan status gizi bayi wilayah kerja Puskesmas Sudiang dengan *Density Based Spatial Clustering of Application with Noise* (DBSCAN)?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini ialah:

- a. Membangun sistem pengelompokkan status gizi bayi wilayah kerja Puskesmas Sudiang dengan *Density Based Spatial Clustering of Application with Noise* (DBSCAN).



- b. Memberikan informasi bagaimana unjuk kerja sistem pengelompokkan status gizi bayi wilayah kerja Puskesmas Sudiang dengan *Density Based Spatial Clustering of Application with Noise* (DBSCAN).

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

- a. Memberikan gambaran tentang pengelompokkan status gizi bayi di wilayah kerja Puskesmas Sudiang Kota Makassar.
- b. Menjadi sumber informasi mengenai pengelompokkan status gizi bayi dengan *Density Based Spatial Clustering of Application with Noise* (DBSCAN).
- c. Menjadi sumber referensi dalam pengembangan bagi penelitian terkait topik *clustering* status gizi bayi.

1.5 Ruang Lingkup

Adapun ruang lingkup dari penelitian ini ialah:

- a. Data yang digunakan adalah data pengukuran bayi yang berasal dari Puskesmas Sudiang Kota Makassar
- b. Variabel-variabel penelitian terdiri dari Provinsi, Kab/Kota, Desa/Kelurahan, Kecamatan, Posyandu, Usia saat ukur, Tinggi, Berat, BB Lahir, TB Lahir, Vit A, Asupan Makanan, Jumlah Keluarga, Pendapatan Orang Tua, dan Riwayat Imunisasi.
- c. Visualisasi hasil *clustering* akan ditampilkan dalam bentuk pemetaan.



BAB II TINJAUAN PUSTAKA

2.1 Status Gizi

Status gizi bayi merupakan indikator penting dalam menilai kesehatan dan kesejahteraan anak. Status gizi didefinisikan sebagai keadaan tubuh sebagai akibat konsumsi makanan dan penggunaan zat-zat gizi (Iskandar, 2022). Penilaian status gizi balita dapat diukur berdasarkan pengukuran antropometri yang terdiri dari variabel umur, berat badan (BB) dan tinggi badan (TB). Umur sangat memegang peranan dalam penentuan status gizi, kesalahan penentuan akan menyebabkan interpretasi status gizi yang salah. Hasil penimbangan berat badan maupun tinggi badan yang akurat, menjadi tidak berarti bila tidak disertai dengan penentuan umur yang tepat. Berat badan merupakan salah satu ukuran yang memberikan gambaran massa jaringan. Berat badan sangat peka terhadap perubahan yang mendadak baik karena penyakit infeksi maupun konsumsi makanan yang menurun. Sedangkan tinggi badan memberikan gambaran fungsi pertumbuhan yang dilihat dari keadaan kurus kering dan kecil pendek. Tinggi badan sangat baik untuk melihat keadaan gizi masa lalu terutama yang berkaitan dengan keadaan berat badan lahir rendah dan kurang gizi pada masa balita (Septikasari, 2018).

Standar Antropometri Anak digunakan untuk menilai atau menentukan status gizi anak. Penilaian status gizi Anak dilakukan dengan membandingkan hasil pengukuran berat badan dan panjang/tinggi badan dengan Standar Antropometri Anak. Menurut Menteri Kesehatan Republik Indonesia (2020), Standar Antropometri Anak didasarkan pada parameter berat badan dan panjang/tinggi badan yang terdiri dari beberapa indeks, meliputi:

- a. Indeks Berat Badan menurut Umur (BB/U), yang menggambarkan berat badan relatif dibandingkan dengan umur anak. Indeks ini digunakan untuk menilai anak dengan berat badan kurang (*underweight*) atau sangat kurang (*severely underweight*), tetapi tidak dapat digunakan untuk mengklasifikasikan anak

atau sangat gemuk. Penting diketahui bahwa seorang anak dengan BB/U rendah, kemungkinan mengalami masalah pertumbuhan, sehingga perlu konfirmasi dengan indeks BB/PB atau BB/TB sebelum diintervensi.



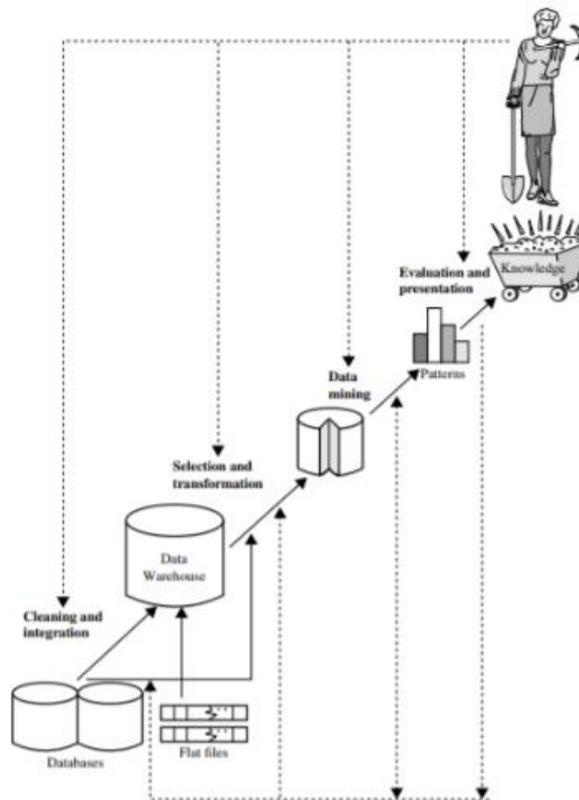
- b. Indeks Panjang Badan menurut Umur atau Tinggi Badan menurut Umur (PB/U atau TB/U), yang menggambarkan pertumbuhan panjang atau tinggi badan anak berdasarkan umurnya. Indeks ini dapat mengidentifikasi anak-anak yang pendek (*stunted*) atau sangat pendek (*severely stunted*), yang disebabkan oleh gizi kurang dalam waktu lama atau sering sakit. Anak-anak yang tergolong tinggi menurut umurnya juga dapat diidentifikasi. Anak-anak dengan tinggi badan di atas normal (tinggi sekali) biasanya disebabkan oleh gangguan endokrin, namun hal ini jarang terjadi di Indonesia.
- c. Indeks Berat Badan menurut Panjang Badan/Tinggi Badan (BB/PB atau BB/TB), yang menggambarkan apakah berat badan anak sesuai terhadap pertumbuhan panjang/tinggi badannya. Indeks ini dapat digunakan untuk mengidentifikasi anak gizi kurang (*wasted*), gizi buruk (*severely wasted*) serta anak yang memiliki risiko gizi lebih (*possible risk of overweight*). Kondisi gizi buruk biasanya disebabkan oleh penyakit dan kekurangan asupan gizi yang baru saja terjadi (akut) maupun yang telah lama terjadi (kronis).

2.2 Data Mining

Data mining adalah proses menemukan pola dan pengetahuan yang menarik dari sejumlah data yang besar (Han dkk., 2011). Menurut Bhatia (2019), *Data Mining* adalah kumpulan teknik untuk penemuan otomatis yang efisien dari pola-pola baru yang sebelumnya tidak diketahui dan valid. Dari definisi ini, hal penting yang dapat diambil adalah:

- a. *Data Mining* adalah proses penemuan otomatis pola yang sebelumnya tidak diketahui dalam jumlah besar volume data yang besar.
- b. Volume data yang besar ini biasanya merupakan data historis dari sebuah organisasi yang dikenal sebagai data gudang.
- c. *Data Mining* berurusan dengan volume data yang besar, dalam *Gigabyte* atau *Terabyte* dan terkadang sebanyak *Zetabyte* data (dalam kasus data besar).





Gambar 1 Tahap *Data Mining* (Han dkk., 2011)

Berdasarkan Gambar 1 Han dkk. (2011), menyebutkan tahapan *Knowledge Discovery from Data* (KDD) atau yang biasa disebut *Data Mining* terbagi atas 7 tahap, yaitu sebagai berikut:

- a. *Data cleaning*: tahap untuk membersihkan data yang hilang, *noise* dan tidak konsisten.
- b. *Data integration*: tahap dimana beberapa sumber data dapat digabungkan.
- c. *Data selection*: tahap untuk data yang relevan dengan analisis diambil dari *database*.
- d. *Data transformation*: tahap untuk data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk penambangan dengan melakukan operasi *summary* atau *aggregation*.
- e. *Data mining*: tahap penting dimana metode cerdas diterapkan untuk mengekstrak pola data.

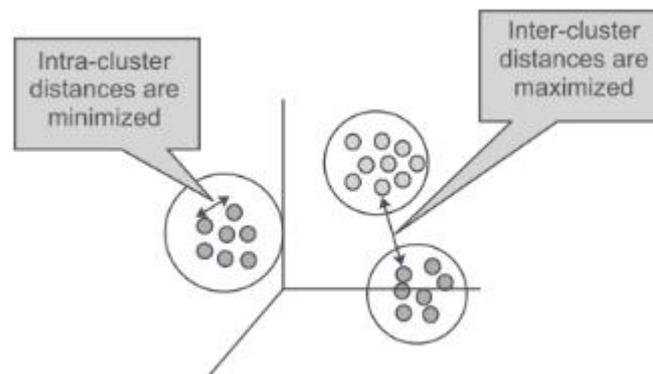


- f. *Pattern evaluation*: tahap untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan ukuran ketertarikannya (*distance/interestingness measure*).
- g. *Knowledge presentation*: tahap dimana teknik gambaran visualisasi dan pengetahuan digunakan untuk menyajikan pengetahuan kepada pengguna.

2.3 Clustering

Clustering adalah proses pengelompokan sekumpulan objek data ke dalam beberapa kelompok atau *cluster* sehingga objek-objek di dalam sebuah *cluster* memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan objek-objek di dalam *cluster* lainnya. Ketidaksamaan dan kemiripan dinilai berdasarkan nilai atribut yang menggambarkan objek dan sering kali melibatkan ukuran jarak (Bhatia, 2019).

Menurut Larose & Larose (2015), *Clustering* berbeda dengan klasifikasi karena tidak ada variabel target untuk *clustering*. Tugas *clustering* tidak mencoba mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel target. Sebaliknya, algoritma clustering mencoba untuk mengelompokkan seluruh kumpulan data ke dalam sub-kelompok atau *cluster* yang relatif seragam, di mana kesamaan catatan dalam *cluster* dimaksimalkan, dan kesamaan dengan catatan di luar *cluster* ini diminimalkan.



Gambar 2 Karakteristik sebuah cluster (Bhatia, 2019).

Bhatia (2019) menyebutkan bahwa *clustering* didefinisikan sebagai ipokan sekumpulan objek yang serupa ke dalam kelas atau *cluster*. Dengan selama analisis *cluster*, data dikelompokkan ke dalam kelas atau *cluster*, data dalam satu *cluster* (*intra-cluster*) memiliki kemiripan yang tinggi



satu sama lain, tetapi memiliki ketidakmiripan yang tinggi dibandingkan dengan objek dalam *cluster* lain (*inter-cluster*) seperti yang ditunjukkan pada gambar 2.

2.4 Principal Component Analysis (PCA)

Teknik reduksi dimensi merupakan hal yang penting dalam banyak aplikasi yang berkaitan dengan *data mining*. Tujuan utama dari teknik reduksi dimensi adalah untuk mentransformasikan data atau fitur dari ruang dimensi yang lebih tinggi ke ruang dimensi yang lebih rendah (Tharwat, 2016). *Principal component analysis* (PCA), yang ditemukan pada tahun 1901 oleh Karl Pearson sebagai teknik reduksi data klasik, menyingkap keterkaitan di antara banyak variabel. PCA mentransformasikan sekumpulan p variabel, X_1, X_2, \dots, X_p menjadi p variabel kombinasi linier PC_1, PC_2, \dots, PC_p (PC untuk komponen utama) sedemikian rupa sehingga sebagian besar informasi dalam sekumpulan variabel asli dapat diwakili dalam sekumpulan variabel baru yang lebih kecil, yang tidak berkorelasi satu sama lain (Ratner, 2017).

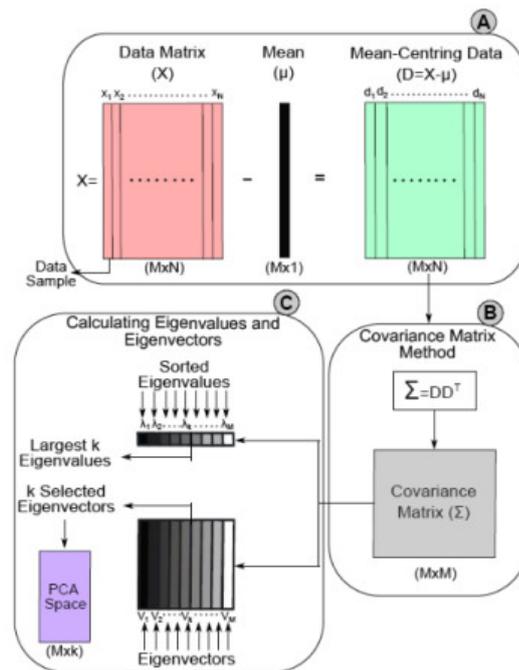
$$\begin{aligned}
 PC_1 &= a_{11} * X_1 + a_{12} * X_2 + \dots + a_{1j} * X_{1j} + \dots + a_{1p} * X_p \\
 PC_2 &= a_{21} * X_1 + a_{22} * X_2 + \dots + a_{2j} * X_{1j} + \dots + a_{2p} * X_p \\
 &\vdots \\
 PC_i &= a_{i1} * X_1 + a_{i2} * X_2 + \dots + a_{ij} * X_{1j} + \dots + a_{ip} * X_p \\
 &\vdots \\
 PC_p &= a_{p1} * X_1 + a_{p2} * X_2 + \dots + a_{pj} * X_{1j} + \dots + a_{pp} * X_p
 \end{aligned} \tag{1}$$

Dimana a_{ij} adalah konstanta yang disebut koefisien PC. X diasumsikan telah distandarisasi. Selain itu, PC dan a_{ij} memiliki banyak sifat aljabar dan interpretasi.

Covariance matrix adalah salah satu metode yang digunakan untuk mendapatkan nilai PCA. Secara umum ada dua langkah yang dilakukan, yaitu menghitung matriks kovarians dari matriks data (X) yang menjadi masukan, dan selanjutnya adalah menghitung nilai eigen dan vektor eigen dari matriks kovarians. Jika kembali pada pengertian dan tujuan PCA, secara matematis, informasi dari data dapat direpresentasikan dengan *covariance matrix* karena dapat mengukur variansi

data) dan juga dapat mengukur korelasi antar satu variabel dengan variabel lain. Maka dari itu *covariance matrix* diperlukan untuk menemukan nilai-nilai λ .





Gambar 3 Ilustrasi proses PCA (Tharwat, 2016)

Menurut Tharwat (2016), berdasarkan gambar 3 berikut adalah langkah-langkah dalam menemukan nilai dan ruang PCA:

1. Temukan *covariance matrix* dari matriks data (X). Persamaan (2) berikut menyatakan struktur dari *covariance matrix*

$$\begin{pmatrix} Var(x_1, x_2) & Cov(x_1, x_2) & \cdots & Cov(x_1, x_M) \\ Cov(x_2, x_1) & Var(x_2, x_2) & \cdots & Cov(x_2, x_M) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_M, x_1) & Cov(x_M, x_2) & \cdots & Var(x_M, x_M) \end{pmatrix} \quad (2)$$

Berdasarkan struktur diatas, dapat dilihat *variance* dan *covariance* data membentuk nilai-nilai dari *covariance matrix*. Matriks kovarians adalah matriks yang simetris ($X = X^T$) dan nilai diagonalnya mewakili variansi dari setiap variabel, sedangkan entri *off-diagonal* mewakili kovarians antara dua variabel yang berbeda. Persamaan (3) merupakan formula untuk menghitung

dari variansi data dan persamaan (4) merupakan formula untuk menghitung nilai dari kovarians data.



$$\sigma^2(x) = \text{Var}(x) = \frac{1}{n-1} \sum_{i=1}^n ((x_i - \mu_i)^2) \quad (3)$$

$$\text{cov}(x_1, x_2) = \frac{1}{n-1} \sum_{i=1}^n [(x_{1i} - \mu(x_1))(x_{2i} - \mu(x_2))] \quad (4)$$

Dimana μ merupakan rata-rata dari variabel yang bersangkutan, dan dapat ditemukan menggunakan persamaan (5) berikut

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (5)$$

Matriks kovarians yang ditemukan akan memiliki nilai positif dan negatif, dimana nilai positif berarti korelasi positif antara kedua variabel, sedangkan nilai negatif menunjukkan korelasi negatif, dan nilai nol menunjukkan bahwa kedua variabel tidak berkorelasi atau independen.

- Setelah matriks kovarians telah ditemukan, selanjutnya hitung nilai eigen dan vektor eigen dari matriks dengan formula (6) berikut:

$$\begin{aligned} Ax &= \lambda x \\ Ax &= \lambda Ix \\ (A - \lambda I)x &= 0 \end{aligned} \quad (6)$$

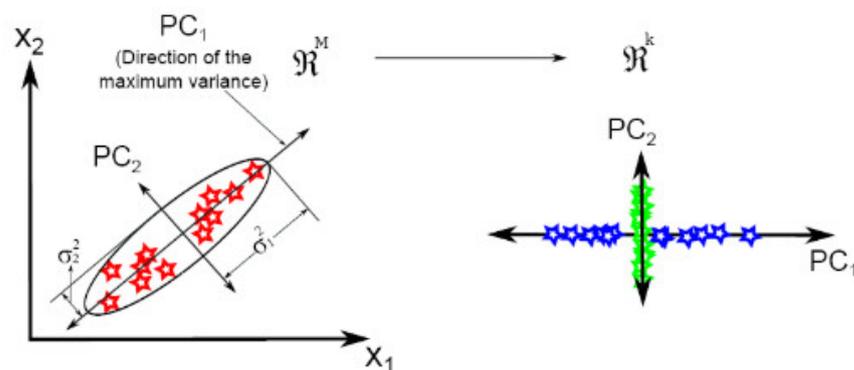
Dimana x dan λ menyatakan vektor eigen dan nilai eigen dari matriks kovarians secara berurut. Nilai eigen merupakan nilai yang skalar, sementara vektor eigen adalah nilai *non-zero*, yang mewakili *principal components* pada *space* baru yang terbentuk sebagai pengganti sumbu dan *space* asli. Untuk melakukan reduksi dimensi, diperlukan pengurutan nilai eigen dari terbesar ke terkecil dan temukan akumulasi nilai eigen yang memiliki variansi minimal 50% variansi yang mewakili informasi dari data asli (Streiner, 1994).

- Setelah nilai eigen dan vektor eigen (*principal components*) ditemukan, selanjutnya adalah menemukan posisi titik-titik data dalam ruang PCA yang dilakukan dengan memproyeksikan matriks *mean-centering data* (matriks selisih data-data asli terhadap rata-rata variabelnya) terhadap vektor eigen yang



telah ditemukan pada tahap sebelumnya dan telah dinormalisasi dan di *transpose*. Sehingga hasil dari proyeksi ini adalah hasil akhir dari penemuan nilai-nilai *principal components* pada PCA.

Pada tahap akhir saat membangun ruang dimensi yang baru dengan PCA, titik-titik data dipusatkan pada nol yang merupakan titik pusat baru pada ruang PCA berdasarkan rata-rata setiap variabel. Seringkali terjadi kondisi meski data matriks masukan semuanya bernilai positif, hasil akhir PCA menunjukkan adanya nilai yang negatif. Nilai negatif pada hasil akhir PCA memiliki 2 penyebab, yang pertama menandakan bahwa nilai data asli lebih kecil atau lebih rendah dari rata-rata variabelnya sehingga matriks *mean-centering* data yang diproyeksikan terhadap *principal components* menjadi negatif. Yang kedua hanya menandakan posisi titik data pada *principal components* di ruang PCA berdasarkan hasil proyeksi *mean-centering data* terhadap vektor eigen (*principal components*). Pada Gambar 4 menunjukkan ilustrasi pembentukan ruang PCA.



Gambar 4 Ilustrasi pembentukan ruang PCA (Tharwat, 2016)

2.5 Euclidian Distance

Data Mining Clustering merupakan teknik atau metode untuk mengelompokkan objek-objek yang memiliki kesamaan/kedekatan ke dalam kluster yang sama, kuncinya adalah kedekatan. Untuk mengetahui kedekatan/kesamaan antar data dalam suatu dataset yang besar, diperlukan suatu an dan perhitungan jarak. Pendekatan standar yang digunakan untuk r jarak adalah *Euclidian Distance Metrics* (EDM). EDM menghitung ara dua titik dalam ruang (*space*). Misalkan ada 2 vektor, $x = (x_1, x_2,$



..., x_n) dan $y = (y_1, y_2, \dots, y_n)$, maka jarak *Euclidian* dapat dihitung dengan persamaan berikut (Han dkk., 2011):

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (7)$$

2.6 Density Based Spatial Clustering of Applications with Noise (DBSCAN)

Metode pengelompokan berbasis kepadatan yang dikenal sebagai DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) adalah sebuah algoritma *clustering* yang melakukan pengelompokan data dengan membentuk *cluster node* berdasarkan estimasi distribusi kepadatan *node* yang sesuai di wilayah tersebut (Bhatia, 2019). Karami & Johansson (2015) menyebut DBSCAN mengelompokkan objek data berdasarkan kepadatannya, yang mana sebuah kluster dianggap sebagai kelompok yang memiliki objek yang padat, dan dipisahkan oleh kelompok yang memiliki kepadatan objek yang rendah atau *noise*. DBSCAN dapat menemukan kluster dengan bentuk yang berubah-ubah serta membedakan *noise*. DBSCAN membutuhkan dua parameter input, yaitu *Eps* (radius kluster) dan *MinPts* (minimum objek data minimum yang diperlukan di dalam kluster).

Menurut Bhatia (2019), untuk memahami cara kerja algoritma DBSCAN, ada beberapa hal yang harus dipahami yang dijelaskan sebagai berikut:

1. *Neighborhood* (Eps)

Neighborhood atau tetangga adalah istilah penting yang digunakan dalam DBSCAN. *Neighborhood* merepresentasikan objek dalam radius tertentu dari objek tipe *centroid*. Hasil *neighborhood* dengan kepadatan tinggi jika sebuah objek mengandung setidaknya *MinPts* (titik minimum) objek di tetangganya. Menurut Karami & Johansson (2015), *Eps-Neighborhood* dari sebuah objek p , dilambangkan dengan $Eps(p)$ dalam sebuah himpunan objek D , didefinisikan oleh $Eps(p) = \{q \in D | distance(p, q) \leq Eps\}$

2. *Core*, *Border*, dan *Outlier*

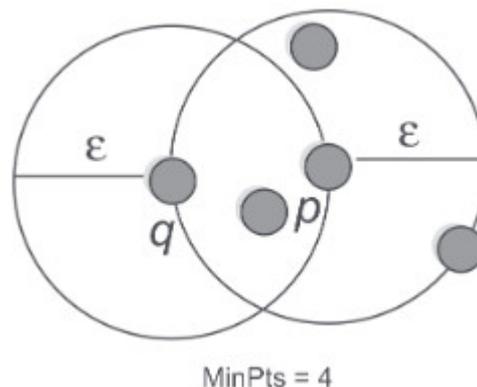
Sebuah titik dikenal sebagai *core* jika memiliki lebih dari jumlah titik yang tukan (*MinPts*) di dalam *neighborhood* (*Eps*). Titik-titik ini harus berada gian dalam kluster. Sebuah *border* adalah sebuah titik jika titik tersebut iliki lebih sedikit dari *MinPts* di dalam *neighborhood* (*Eps*), tetapi berada



di dalam *neighborhood* dari *core*. Sebuah titik adalah titik *noise* atau titik *outlier* jika titik tersebut bukan merupakan *core* maupun *border*.

3. *Density reachability*

Definisi kluster dari DBSCAN ditentukan berdasarkan konsep *density-reachability*. Secara umum, sebuah titik q secara langsung *density-reachable* dari sebuah titik p jika titik tersebut tidak lebih jauh dari jarak ϵ (Eps) yang diberikan. Namun, kita dapat menganggap bahwa p dan q termasuk dalam *cluster* yang sama yang sama jika titik p dikelilingi oleh banyak titik.



Gambar 5 Konsep dari *density-reachability* (Bhatia, 2019)

Pada Gambar 5 dijelaskan tentang konsep *density-reachability*. Ada beberapa hal yang dapat dijelaskan pada gambar tersebut.

- jika p adalah *core* dan titik q berada dalam jangkauan ϵ dari p maka kita dapat mengatakan bahwa titik q adalah *directly density-reachable* dari titik p .
- q adalah *directly density-reachable* dari *core* p .
- q adalah bukan sebuah *core* karena hanya memiliki 3 titik didalam jangkauan *eps-neighborhood* nya sedangkan MinPts adalah 4 titik.
- p tidak termasuk *directly density-reachable* dari q .

2.6.1 Langkah-langkah Algoritma DBSCAN



nurut Han dkk. (2011), Proses DBSCAN dapat dibagi menjadi beberapa sebagai berikut:

1) objek dalam data set D awalnya ditandai sebagai "*unvisited*". Pilih secara acak sebuah objek p yang belum dikunjungi. Tandai objek p sebagai "*visited*".

2. Periksa apakah *Eps-neighborhood* (lingkungan dengan jarak eps) dari objek p berisi setidaknya $MinPts$ objek. Jika *Eps-neighborhood* tidak memenuhi syarat $MinPts$, maka objek p dianggap sebagai *noise*.
3. Jika *Eps-neighborhood* dari objek p memenuhi syarat $MinPts$, maka buatlah sebuah klaster baru C untuk objek p . Tambahkan semua objek dalam *Eps-neighborhood* dari p ke dalam himpunan kandidat N .
4. Iterasikan melalui objek-objek dalam himpunan kandidat N . Untuk setiap objek p' dalam N yang belum dikunjungi: Tandai objek p' sebagai "*visited*". Periksa *Eps-neighborhood* dari objek p' . Jika *Eps-neighborhood* dari p' memenuhi syarat $MinPts$, tambahkan semua objek dalam *Eps-neighborhood* dari p' ke dalam himpunan kandidat N . Tambahkan objek p' ke dalam klaster C jika objek p' tidak termasuk dalam klaster manapun.
5. Teruskan proses ekspansi klaster C hingga tidak ada lagi objek yang dapat ditambahkan ke dalamnya. Pada titik ini, klaster C dianggap selesai dan hasil keluaran.

Adapun *pseudocode* dari algoritma DBSCAN diberikan sebagai Gambar 6 berikut (Han dkk., 2011):

```

Algorithm: DBSCAN: a density-based clustering algorithm.
Input:
  ■  $D$ : a data set containing  $n$  objects,
  ■  $\epsilon$ : the radius parameter, and
  ■  $MinPts$ : the neighborhood density threshold.
Output: A set of density-based clusters.
Method:
(1) mark all objects as unvisited;
(2) do
(3)   randomly select an unvisited object  $p$ ;
(4)   mark  $p$  as visited;
(5)   if the  $\epsilon$ -neighborhood of  $p$  has at least  $MinPts$  objects
(6)     create a new cluster  $C$ , and add  $p$  to  $C$ ;
(7)     let  $N$  be the set of objects in the  $\epsilon$ -neighborhood of  $p$ ;
(8)     for each point  $p'$  in  $N$ 
(9)       if  $p'$  is unvisited
(10)        mark  $p'$  as visited;
(11)        if the  $\epsilon$ -neighborhood of  $p'$  has at least  $MinPts$  points,
            add those points to  $N$ ;
(12)        if  $p'$  is not yet a member of any cluster, add  $p'$  to  $C$ ;
(13)     end for
(14)   output  $C$ ;
(15) else mark  $p$  as noise;
(16) until no object is unvisited;

```

Gambar 6 *Pseudocode* algoritma DBSCAN (Han dkk., 2011)



2.6.2 Menentukan Parameter pada DBSCAN

Dalam pembangunan sebuah cluster pada algoritma DBSCAN dibutuhkan 2 parameter, yaitu MinPts dan Eps. Untuk menemukan kedua parameter tersebut, parameter tersebut dapat disesuaikan dengan kebutuhan kasus, namun juga dapat ditemukan dengan pendekatan heuristik. Berikut adalah aturan yang dapat digunakan untuk menentukan nilai MinPts dan Eps (Ester dkk., 1996):

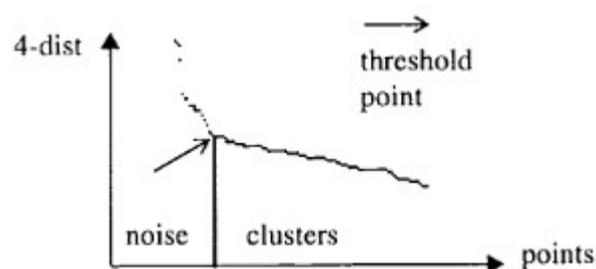
1. *Minimum Points* (MinPts)

Sebagai aturan praktis, MinPts dapat diperoleh dengan cara sebagai berikut:

- a. MinPts sama dengan jumlah dimensi D dalam kumpulan data, karena $\text{MinPts} \geq D + 1$.
- b. MinPts harus dipilih minimal 3. Namun, nilai yang lebih besar biasanya lebih baik untuk kumpulan data dengan *noise* dan akan menghasilkan kluster yang lebih signifikan.
- c. Sebagai aturan praktis, $\text{MinPts} = 2 * \text{dimensi}$ dapat digunakan (Sander & Ester, 1998).

2. *Epsilon* (Eps)

Nilai Eps dapat dipilih dengan menggunakan *ordered k-dist-graph*, memplotkan jarak ke tetangga terdekat k yang diurutkan dari nilai terbesar hingga terkecil. Nilai Eps yang baik adalah ketika plot ini menunjukkan "siku-siku" (Schubert dkk., 2017). Jika Eps dipilih terlalu kecil, sebagian besar data tidak akan terkluster, sedangkan untuk nilai Eps yang terlalu besar, kluster-kluster akan menyatu dan sebagian besar objek berada dalam satu kluster. Pada DBSCAN, k dapat merujuk pada nilai MinPts, sebab dapat menemukan jarak-jarak terjauh (Eps) yang dapat memenuhi target MinPts (Ester dkk., 1996).



Gambar 7 Ilustrasi *ordered k-dist-graph* (Ester dkk., 1996)

la Gambar 7 dijelaskan sebuah ilustrasi penggunaan *ordered k-dist-graph*. Kita memilih sebuah titik sembarang p , mengatur parameter Eps ke k -



$dist(p)$ dan mengatur parameter MinPts ke k , semua titik dengan nilai $k-dist$ yang sama atau lebih kecil akan menjadi titik inti. Jika kita dapat menemukan titik ambang dengan nilai $k-dist$ maksimal di kluster D yang "paling tipis", kita akan mendapatkan nilai parameter yang diinginkan. Titik ambang batas adalah titik pertama di "lembah" pertama dari $k-dist-graph$ yang telah diurutkan. Semua titik dengan nilai $k-dist$ yang lebih tinggi (sebelah kiri ambang batas) dianggap sebagai *noise*, semua titik lainnya (sebelah kanan ambang batas) dimasukkan ke dalam suatu kluster (Ester dkk., 1996).

2.6.3 Kelebihan dan Kekurangan Algoritma DBSCAN

Bhatia (2019) menjelaskan tentang beberapa kelebihan Algoritma DBSCAN dibanding dengan algoritma lainnya.

1. Tidak perlu menentukan jumlah kluster dalam data di awal dalam kasus algoritma DBSCAN.
2. Hanya membutuhkan dua parameter dan tidak bergantung pada urutan titik-titik dalam *database*.
3. Algoritma ini dapat mengidentifikasi kluster dengan bentuk yang berubah-ubah. Ia juga dapat mengidentifikasi kluster sepenuhnya dikelilingi oleh kluster yang berbeda.
4. DBSCAN kuat terhadap *outlier* dan memiliki pengertian tentang *noise*.

Selain beberapa kelebihan tersebut, algoritma DBSCAN juga memiliki beberapa kelemahan. Kelemahan Algoritma DBSCAN dijabarkan oleh Bhatia (2019) sebagai berikut.

1. Algoritma DBSCAN sensitif terhadap parameter, yaitu, sulit untuk mengidentifikasi yang benar set parameter.
2. Kualitas algoritma DBSCAN bergantung pada ukuran jarak seperti *Neighborhood* (Eps) dan MinPts.
3. Algoritma DBSCAN tidak dapat mengelompokkan *dataset* secara akurat

dan kepadatan yang bervariasi atau besar perbedaan dalam kepadatan.



2.7 Silhouette Coefficient

Silhouette Coefficient (SC) adalah metode yang digunakan untuk mengukur seberapa baik penetapan *cluster* untuk titik tertentu. Nilai *silhouette* positif menunjukkan bahwa penetapan nilainya baik, dengan nilai yang lebih tinggi lebih baik daripada nilai yang lebih rendah. Nilai *silhouette* yang mendekati nol dianggap sebagai penugasan yang lemah. Nilai *silhouette* negatif dianggap sebagai kesalahan klasifikasi, karena penugasan ke kluster terdekat berikutnya akan lebih baik (Larose & Larose, 2015). Nilai *silhouette* dari setiap data dapat didefinisikan sebagai berikut:

$$Silhouette_i = S_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (8)$$

Dimana a_i adalah jarak antara nilai data dengan pusat kluster, dan b_i adalah jarak antara nilai data dengan pusat kluster terdekat berikutnya. Menurut Kaufman & Rousseeuw (1990), ukuran *Silhouette Coefficient* dapat dijelaskan seperti pada tabel berikut:

Tabel 1 Interpretasi *Silhouette Coefficient* (Kaufman & Rousseeuw, 1990)

Silhouette Coefficient	Interpretasi
0,71 – 1,00	Struktur yang kuat telah ditemukan
0,51 – 0,70	Struktur yang masuk akal telah ditemukan
0,26 – 0,50	Strukturnya lemah dan bisa jadi rekayasa; silakan coba metode tambahan.
$\leq 0,25$	Tidak ada struktur substansial yang ditemukan

Berdasarkan Tabel 1, *Silhouette Coefficient* yang mendekati nilai 1 menunjukkan struktur yang sangat jelas dan nilai *Silhouette Coefficient* yang rendah mengindikasikan bahwa lebih baik menerapkan metode analisis data alternatif.



2.8 Sum of Squared Error (SSE)

Sum of Squared Error (SSE) merupakan salah satu cara untuk mengukur *clustering* dengan menggunakan teknik statistik yang mampu mencari apakah objek cocok pada satu klaster. Adapun perhitungan dari SSE ialah sebagai berikut:

$$SSE = \sum_{i=1}^n (d_i - c_i)^2 \quad (9)$$

Dimana:

$$c_i = \frac{\sum_{i=1}^n x_i}{n} \quad (10)$$

Keterangan:

SSE = nilai kuadrat selisih antara koordinat *centroid* ke setiap data

n = jumlah data

d_i = nilai data ke-i

c_i = nilai *centroid* klaster ke-i

Bila objek sangat cocok dengan klaster tersebut maka nilai SSE adalah nol atau berarti tidak ada *error* atau sangat cocok. Namun hal itu jarang terjadi, oleh karena itu, *clustering* yang baik adalah yang memiliki nilai SSE serendah mungkin. Semakin rendah nilai SSE maka semakin sama. SSE yang tinggi maka memiliki derajat perbedaan antara objek dan klaster yang dituju (Shofiani, 2017).

2.9 Pemetaan

Pemetaan adalah pengelompokkan suatu kumpulan wilayah yang berkaitan dengan beberapa letak geografis wilayah yang meliputi dataran tinggi, pegunungan, sumber daya dan potensi penduduk yang berpengaruh terhadap sosial kultural yang memiliki ciri khas khusus (Rusdiyanto, 2017). Pengertian lain tentang pemetaan yaitu sebuah tahapan yang harus dilakukan dalam pembuatan peta. Langkah awal

kukan dalam pembuatan data, dilanjutkan dengan pengolahan data, dan dalam bentuk peta (Mudhari, 2018).



Jadi, dari dua definisi di atas dan disesuaikan dengan penelitian ini maka pemetaan merupakan proses pengumpulan data untuk dijadikan sebagai langkah awal dalam pembuatan peta, dengan menggambarkan penyebaran kondisi alamiah tertentu secara meruang, memindahkan keadaan sesungguhnya kedalam peta dasar, yang dinyatakan dengan penggunaan skala peta.

