

**PEMETAAN PENYEBARAN PENYAKIT TUBERKULOSIS  
MENGUNAKAN METODE CLUSTERING DBSCAN DAN  
OPTICS PADA PROVINSI ACEH**

**TUGAS AKHIR**

**Diajukan oleh:**

**HADI ALBAR  
210705035**

**Mahasiswa Fakultas Sains dan Teknologi  
Program Studi Teknologi Informasi**



**FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI AR-RANIRY  
BANDA ACEH  
2026 M / 1447 H**

## LEMBAR PENGESAHAN

### PERBANDINGAN METODE DBSCAN DAN OPTICS DALAM PEMETAAN PENYEBARAN PENYAKIT TUBERKULOSIS DI PROVINSI ACEH TUGAS AKHIR

Telah Diuji Oleh Panitia Ujian Munaqasyah Tugas Akhir  
Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh dan Dinyatakan Lulus  
Serta Diterima Sebagai Salah Satu Beban Studi Program Sarjana (S-1)  
Dalam Program Studi Teknologi Informasi

Pada Hari/Tanggal: Selasa, 27 Januari 2026

8 Sya'ban 1447 H

Di Darussalam, Banda Aceh

Panitia Ujian Munaqasyah Tugas Akhir:

Ketua,



Dr. Hendri Ahmadian, S.Si., M.I.M

NIP. 198301042014031002

Sekretaris,



Mursyidin, M.T

NIP. 198204052023211020

Penguji I,



Malahayati, M.T

NIP.198301272015032003

Penguji II,



Baihaqi, M.T

NIP. 198802212022031001

Mengetahui:

Dekan Fakultas Sains dan Teknologi

UIN Ar-Raniry Banda Aceh,



Prof. Dr. Ir. Muhammad Dirhamsyah, M.T., IPU

NIP.196210021988111001

## LEMBAR PERSETUJUAN

### PERBANDINGAN METODE DBSCAN DAN OPTICS DALAM PEMETAAN PENYEBARAN PENYAKIT TUBERKULOSIS DI PROVINSI ACEH

#### TUGAS AKHIR

Diajukan kepada Fakultas Sains dan Teknologi  
Universitas Islam Negeri Ar-Raniry Banda Aceh  
Sebagai Salah Satu Beban Studi Memperoleh Gelar Sarjana (S1)  
Dalam Prodi Teknologi Informasi

Oleh:

**HADI ALBAR**


**NIM. 210705035**


**Mahasiswa Fakultas Sains dan Teknologi  
Program Studi Teknologi Informasi**

Disetujui Untuk Munaqasyah Oleh:

Pembimbing I,

Pembimbing II,

  
**Dr. Hendri Ahmadian, S.Si., M.I.M.**

  
**Mursyidin, M.T.**

**NIP.198301042014031002**

**NIP. 198204052023211020**

Mengetahui,  
Ketua Program Studi Teknologi Informasi



**Malahayati, M.T**  
**NIP.198301272015032003**

## LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Hadi Albar  
NIM : 210705035  
Program Studi : Teknologi Informasi  
Fakultas : Sains dan Teknologi  
Judul : PERBANDINGAN METODE DBSCAN DAN OPTICS  
DALAM PEMETAAN PENYEBARAN PENYAKIT  
TUBERKULOSIS DI PROVINSI ACEH

Dengan ini menyatakan bahwa dalam penulisan tugas akhir ini, saya:

1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan;
2. Tidak melakukan plagiasi terhadap naskah karya orang lain;
3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa izin pemilik karya;
4. Tidak memanipulasi dan memalsukan data;
5. Mengerjakan sendiri karya ini dan mampu bertanggungjawab atas karya ini.

Bila dikemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggungjawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenai sanksi berdasarkan aturan yang berlaku di Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh. Demikian pernyataan ini saya buat dengan sesungguhnya dan tanpa paksaan dari pihak manapun.

Banda Aceh, 26 Januari 2026



Yang Menyatakan

  
Hadi Albar

## KATA PENGANTAR

*Bismillahirrahmanirrahim.*

Dengan menyebut nama Allah Yang Maha Pengasih lagi Maha Penyayang. Puji syukur penulis panjatkan kehadiran Allah SWT yang atas berkat, rahmat, dan karunia-Nya telah memberikan kekuatan dan kemudahan, sehingga penulis berhasil merampungkan Tugas Akhir yang berjudul **“PERBANDINGAN METODE DBSCAN DAN OPTICS DALAM PEMETAAN PENYEBARAN PENYAKIT TUBERKULOSIS DI PROVINSI ACEH”**. Shalawat dan salam semoga senantiasa terlimpahkan kepada teladan umat manusia, Nabi Muhammad SAW, beserta keluarga, sahabat, dan para pengikutnya hingga akhir zaman.

Penulisan Tugas Akhir ini diajukan untuk memenuhi salah satu syarat akademis dalam memperoleh gelar Sarjana (S1) pada Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry.

Penulis menyadari bahwa perjalanan dan penyelesaian studi ini tidak lepas dari bimbingan, dukungan, dan doa dari banyak pihak. Untuk itu, dengan penuh rasa hormat dan terima kasih, penulis ingin menghaturkan penghargaan yang tulus kepada:

1. Ayahanda Dahlan Hamid dan ibunda Sulmiati, pilar utama kekuatan dan sumber inspirasi. Terima kasih atas limpahan kasih sayang yang tiada henti, doa yang tak pernah terputus, serta dukungan moril dan materiel yang menjadi fondasi bagi penulis dalam menempuh pendidikan hingga titik ini. Semoga Allah SWT senantiasa melimpahkan karunia kesehatan dan kebahagiaan untuk ayah dan bunda.
2. Ibu Malahayati, M.T., selaku Ketua Program Studi Teknologi Informasi. Terima kasih yang sebesar-besarnya atas kesabaran, waktu, serta bimbingan akademis yang sangat berharga. Arahan, masukan, dan diskusi yang diberikan Beliau sangat membantu dalam mengarahkan penelitian ini.

3. Bapak Dr. Hendri Ahmadian, S.Si., M.I.M., selaku Pembimbing I, yang telah banyak meluangkan waktu untuk memberikan bimbingan, saran, dan diskusi yang mencerahkan sehingga penelitian ini dapat terselesaikan dengan lebih baik.
4. Bapak Mursyidin, M.T., selaku Pembimbing II, yang dengan sabar memberikan arahan, masukan, dan dukungan intelektual sejak awal proses penelitian hingga selesai.
5. Ibu Cut Ida Rahmadiana, S.Si., atas segala bantuan, kemudahan, dan kesabaran dalam proses pengurusan administrasi akademik yang sangat menunjang kelancaran studi penulis.
6. Bapak Prof. Dr. Ir. M. Dirhamsyah, M.T., IPU, selaku Dekan Fakultas Sains dan Teknologi UIN Ar-Raniry, atas segala dukungan dan fasilitas yang diberikan selama masa studi
7. Abang tercinta, Miftahul Razaq, atas doa, perhatian, dan semangat yang tidak pernah henti, serta peran besarnya dalam memberikan dukungan moral selama penyusunan Tugas Akhir ini.
8. Seluruh Bapak dan Ibu Dosen di lingkungan Program Studi Teknologi Informasi yang telah membagikan ilmu pengetahuan dan wawasan yang tak ternilai.
9. Sahabat dan rekan seperjuangan: Fahmi, Azuwan, Haris, Yudhi, Arief, Ammar, Aidil, Amir yang telah menjadi teman berbagi dalam suka, duka, dan segala proses yang dijalani. Terima kasih atas semangat dan dukungan yang tulus.
10. Kawan-kawan grup lari WGS, Febri, Ipa, Popon, dan Una terima kasih atas kebersamaan, dukungan emosional, dan energi positif yang senantiasa diberikan. Kegiatan berlari bersama WGS tidak hanya membantu menjaga kesehatan fisik, tetapi juga berperan penting dalam menjaga kesehatan mental penulis, terutama di tengah tekanan dan dinamika penyusunan Tugas Akhir. Kebersamaan ini menjadi ruang pelepas penat, penguat semangat, dan sumber keseimbangan yang sangat berarti hingga tahap penyelesaian skripsi ini.

11. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan kontribusi dalam penyelesaian Tugas Akhir ini.

Penulis mengakui bahwa karya ini masih memiliki banyak kekurangan. Oleh karena itu, segala bentuk saran dan kritik yang bersifat membangun akan penulis terima dengan lapang dada demi penyempurnaan di masa yang akan datang.

Akhir kata, semoga Tugas Akhir ini dapat bermanfaat bagi penulis, pembaca, dan pengembangan ilmu pengetahuan. Semoga Allah SWT mencatatnya sebagai amal kebaikan.

*Aamiin yaa Rabbal'aalamiin.*

Banda Aceh, 26 Januari 2026

Penulis,

Hadi Albar



## ABSTRAK

Nama : Hadi Albar  
NIM : 210705035  
Program Studi : Teknologi Informasi  
Fakultas : Sains dan Teknologi  
Judul Tugas Akhir : Perbandingan Metode Dbscan Dan  
Optics Dalam Pemetaan Penyebaran  
Penyakit Tuberkulosis Di Provinsi Aceh  
Tanggal Sidang : 27 Januari 2026  
Jumlah Halaman : 79  
Pembimbing I : Dr. Hendri Ahmadian, S.Si., M.I.M.  
Pembimbing II : Mursyidin, M.T.

Persebaran tuberkulosis (TB) yang tidak merata dipengaruhi oleh perbedaan kondisi geografis, sosial, dan ekonomi, sehingga pemetaan berbasis data spasial diperlukan untuk melihat pola sebaran secara lebih jelas. Penelitian ini bertujuan membandingkan efektivitas metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dan *Ordering Points To Identify the Clustering Structure* (OPTICS) dalam memetakan pola penyebaran TB di Provinsi Aceh. Penelitian ini bersifat deskriptif kuantitatif dengan pendekatan data mining menggunakan data sekunder dari E-Profil Kesehatan Aceh dan Global Tuberculosis Report 2023. Variabel yang dianalisis meliputi koordinat geografis (latitude dan longitude) serta jumlah kasus TB tahun 2019–2023. Tahapan analisis mencakup data cleaning, standarisasi, penentuan parameter  $\epsilon$  DBSCAN melalui K-Nearest Neighbors Distance Plot dengan titik elbow pada kisaran 0,8–1,0 ( $k=2$ ), penerapan DBSCAN dan OPTICS, evaluasi kualitas kluster menggunakan Silhouette, serta visualisasi dalam peta. Hasil DBSCAN membentuk dua kluster, dengan sebagian besar kabupaten/kota berada pada kluster dominan, sedangkan Aceh Tengah, Gayo Lues, dan Nagan Raya berada pada kluster yang menunjukkan pola sebaran yang menyimpang. OPTICS menghasilkan tiga kluster

dan mengidentifikasi satu kluster penyimpang (outlier), dengan Simeulue sebagai wilayah yang memiliki pola berbeda dari mayoritas. Evaluasi menunjukkan OPTICS memperoleh Silhouette Score rata-rata 0,409 ( $\approx 0,41$ ) dan lebih adaptif pada variasi kepadatan, sedangkan DBSCAN cenderung menghasilkan lebih banyak data yang terklasifikasi sebagai *noise*.

**Kata kunci:** tuberkulosis, pemetaan spasial, clustering, DBSCAN, OPTICS, Aceh.



## ABSTRACT

Name : Hadi Albar  
Student ID : 210705035  
Major : Information Technology  
Fakulty : Science and Technology  
Final Project Title : Comparison of Dbscan and Optics  
Methods in Mapping the Spread of  
Tuberculosis in Aceh Province  
Thesis Defence Date : 27 January 2026  
Number Of Pages : 79  
Supervisor I : Dr. Hendri Ahmadian, S.Si., M.I.M.  
Supervisor II : Mursyidin, M.T.

The uneven distribution of tuberculosis (TB) is influenced by differences in geographical, social, and economic conditions, so spatial data-based mapping is needed to see the distribution patterns more clearly. This study aims to compare the effectiveness of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Ordering Points To Identify the Clustering Structure (OPTICS) methods in mapping TB distribution patterns in Aceh Province. This study is descriptive and quantitative in nature, using a data mining approach with secondary data from the Aceh Health E-Profile and the Global Tuberculosis Report 2023. The variables analyzed include geographic coordinates (latitude and longitude) and the number of TB cases from 2019 to 2023. The analysis stages included data cleaning, standardization, determination of the DBSCAN  $\epsilon$  parameter through K-Nearest Neighbors Distance Plot with an elbow point in the range of 0.8–1.0 ( $k=2$ ), application of DBSCAN and OPTICS, evaluation of cluster quality using Silhouette, and visualization in a map. The DBSCAN results formed two clusters, with most districts/cities in the dominant cluster, while Central Aceh, Gayo Lues, and Nagan Raya were in a cluster showing a deviant distribution pattern. OPTICS produced three clusters and identified one outlier cluster, with Simeulue as the region with a pattern different from the majority. The evaluation shows that OPTICS obtained an average Silhouette Score of 0.409 ( $\approx 0.41$ ) and was more adaptive to density variations, while DBSCAN tended to produce more data classified as noise.

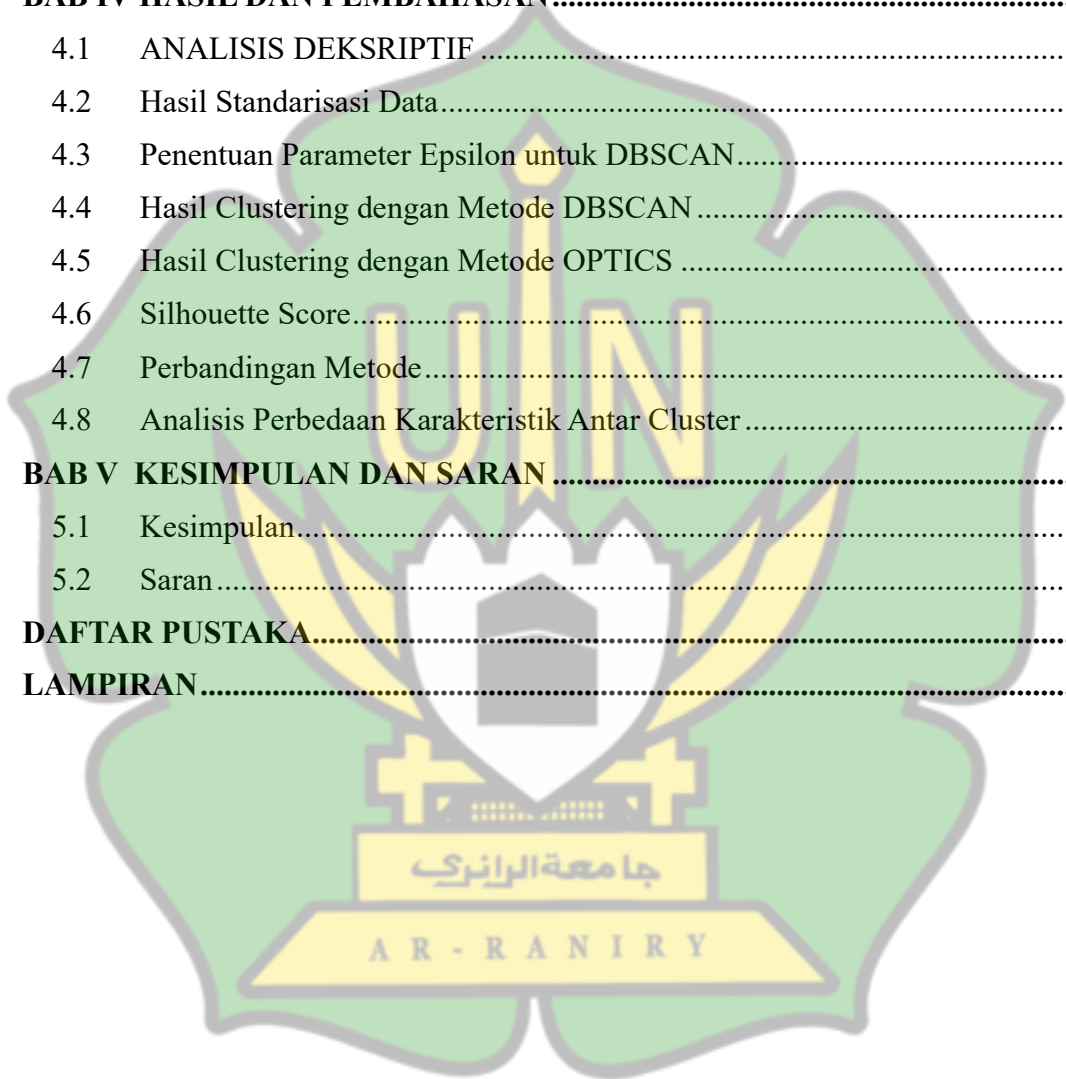
**Keywords:** tuberculosis, spatial mapping, clustering, DBSCAN, OPTICS, Aceh.



# DAFTAR ISI

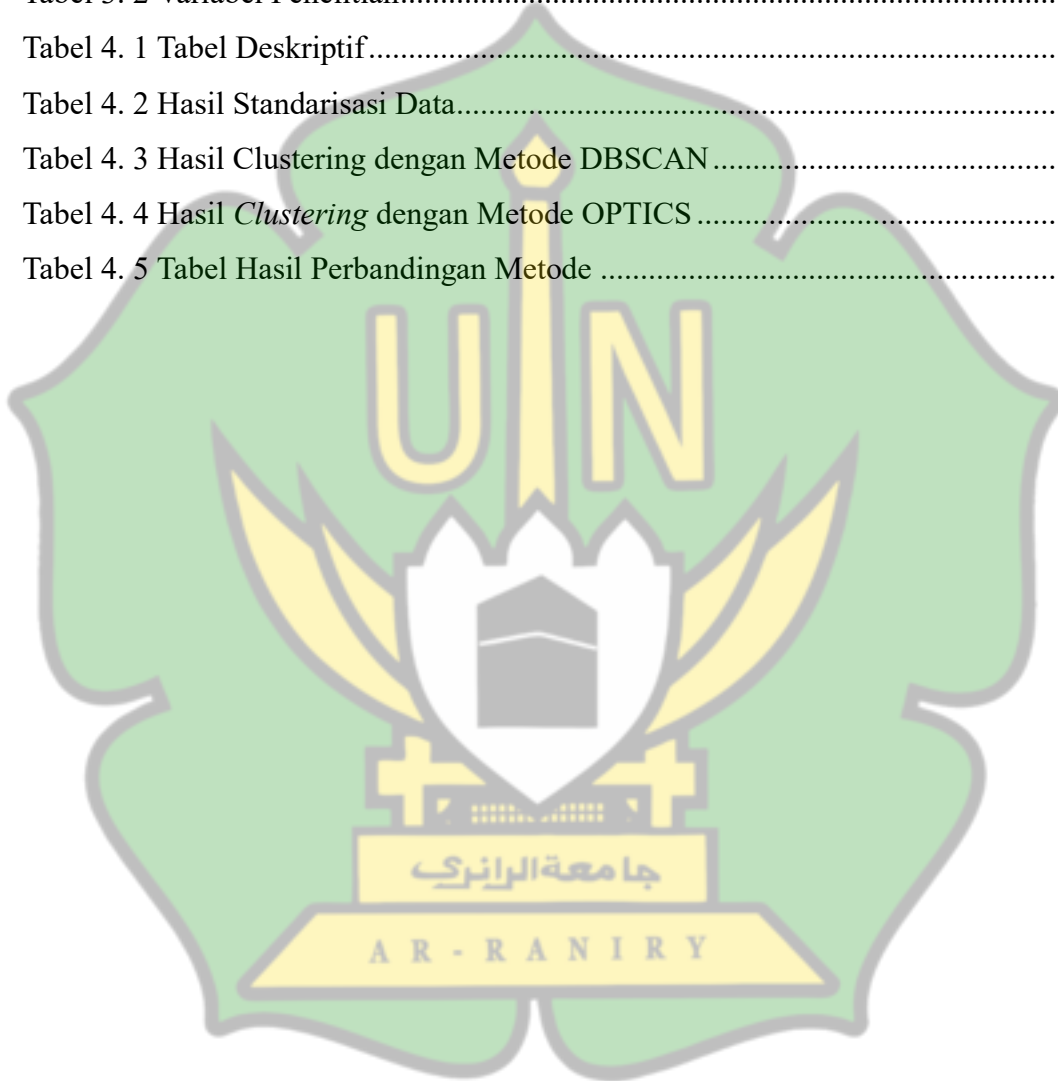
<b>LEMBAR PERSETUJUAN</b> .....	<b>ii</b>
<b>LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR</b> .....	<b>iv</b>
<b>KATA PENGANTAR</b> .....	<b>v</b>
<b>ABSTRAK</b> .....	<b>viii</b>
<b>ABSTRACT</b> .....	<b>x</b>
<b>DAFTAR ISI</b> .....	<b>xii</b>
<b>DAFTAR TABEL</b> .....	<b>xiv</b>
<b>DAFTAR GAMBAR</b> .....	<b>xv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian.....	3
1.4. Batasan Masalah.....	4
1.5. Manfaat Penelitian.....	4
<b>BAB II. TINJAUAN KEPUSTAKAAN</b> .....	<b>5</b>
2.1 Penelitian Terdahulu.....	5
2.2 Tuberkulosis (TB).....	10
2.3 Data Mining.....	11
2.4 Clustering.....	14
2.5 Algoritma DBSCAN ( <i>Density-Based Spatial Clustering of Applications with Noise</i> ).....	15
2.6 Algoritma OPTICS ( <i>Ordering Points To Identify the Clustering Structure</i> ).....	18
2.7 Pengujian Indeks Silhouette.....	20
<b>BAB III METODE PENELITIAN</b> .....	<b>21</b>
3.1. Jenis Penelitian .....	21
3.2. Tahapan Penelitian.....	21
3.2.1. Pengumpulan Data .....	22
3.2.2. Data Cleaning.....	23
3.2.3. Mencari Nilai Epsilon .....	24

3.2.4.	Clustering dengan Metode DBSCAN .....	24
3.2.6.	Pengujian Validitas Indeks Silhouette .....	25
3.3.	Data Dan Variabel Penelitian.....	25
3.4.	Variabel Penelitiann.....	26
3.5.	Data dan Sumber Dataaa .....	26
<b>BAB IV</b>	<b>HASIL DAN PEMBAHASAN.....</b>	<b>27</b>
4.1	ANALISIS DEKSRIPTIF .....	27
4.2	Hasil Standarisasi Data.....	30
4.3	Penentuan Parameter Epsilon untuk DBSCAN.....	31
4.4	Hasil Clustering dengan Metode DBSCAN .....	33
4.5	Hasil Clustering dengan Metode OPTICS .....	36
4.6	Silhouette Score.....	42
4.7	Perbandingan Metode.....	43
4.8	Analisis Perbedaan Karakteristik Antar Cluster .....	44
<b>BAB V</b>	<b>KESIMPULAN DAN SARAN .....</b>	<b>45</b>
5.1	Kesimpulan.....	45
5.2	Saran.....	46
<b>DAFTAR PUSTAKA.....</b>		<b>47</b>
<b>LAMPIRAN.....</b>		<b>56</b>



## DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu.....	5
Tabel 3. 1 Data TB Aceh 2019-2023.....	23
Tabel 3. 2 Variabel Penelitian.....	26
Tabel 4. 1 Tabel Deskriptif.....	27
Tabel 4. 2 Hasil Standarisasi Data.....	30
Tabel 4. 3 Hasil Clustering dengan Metode DBSCAN.....	35
Tabel 4. 4 Hasil <i>Clustering</i> dengan Metode OPTICS.....	40
Tabel 4. 5 Tabel Hasil Perbandingan Metode.....	43



## DAFTAR GAMBAR

Gambar 2. 1 Diagram Proses Data Mining .....	11
Gambar 2. 2 Ilustrasi Directly density-reachable.....	16
Gambar 2. 3 Ilustrasi Density-reachable .....	16
Gambar 2. 4 Ilustrasi <i>Density-connected</i> .....	17
Gambar 3. 1 Diagram Alir Penelitian.....	22
Gambar 4. 1 Boxplot kasus TB 2019 - 2023.....	29
Gambar 4. 2 <i>K-Nearest Neighbors Distance Plot</i> .....	32
Gambar 4. 3 Hasil Clustering dengan Metode DBSCAN.....	33
Gambar 4. 4 Peta Clustering dengan Metode DBSCAN .....	35
Gambar 4.5 Hasil Clustering dengan Metode OPTICS .....	37
Gambar 4. 6 Peta Clustering dengan Metode OPTICS.....	39
Gambar 4. 7 Gambar Perbedaan Karakteristik Antar Cluster .....	44



# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Kejadian penyebaran penyakit TB (Tuberkulosis) yang telah dinyatakan oleh WHO (World Health Organization) sebagai sebuah persoalan yang besar bagi seluruh negara di dunia. TB sendiri merupakan penyakit infeksi menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis* (Kemenkes RI, 2023). Gejala yang sering muncul pada penderita TB diantaranya yaitu demam, batuk selama 2 minggu atau lebih, disertai berkeringat pada malam hari, sesak nafas, nyeri dada, lelah, penurunan berat badan, dan malaise (Peraturan Menteri Kesehatan Republik Indonesia Nomor 67 Tahun 2016, 2016). Berdasarkan Global tuberculosis report tahun 2024 jumlah kasus TB di dunia saat ini mencapai 8,2 juta jiwa, dimana India 26%, Indonesia 10%, dan Tiongkok 6,8% menjadi negara dengan jumlah kasus TB terbanyak di dunia.

Menurut laporan dari Global tuberculosis report tahun 2024, Indonesia berada di peringkat kedua di seluruh dunia sebanyak 1.060.000, dan kematian akibat TB sebanyak 134.000, per 1 Maret 2024 menunjukkan sekitar 821.200 kasus TB baru. Tiga daerah di Indonesia memiliki tingkat kasus TB tertinggi ialah Sumatera (33%), Jawa dan Bali (23%), dan Indonesia Timur (44%) (Amalia, 2020). Penyakit tuberkulosis ini adalah penyebab kematian ketiga terbesar di semua kelompok usia, setelah penyakit jantung dan penyakit saluran pernapasan, dan juga penyebab kematian tertinggi di antara penyakit infeksi. Jumlah kematian akibat TB di Indonesia mencapai 93.000 per tahun pada tahun 2022, atau sekitar 11 kematian per jam, menurut data Kementerian Kesehatan RI (Kemenkes RI, 2023).

Di Aceh, pada tahun 2021 tercatat 7.565 kasus TB, yang kemudian meningkat menjadi 12.286 kasus di tahun 2022, dan hingga 2023 jumlahnya telah bertambah lagi menjadi 13.057 kasus (Profil Kesehatan Aceh, 2023). Penyebaran TB ini dipengaruhi oleh variasi kondisi wilayah dan waktu (Kodariah et al., 2023). Tingkat kerentanan masyarakat terhadap TB juga bervariasi, tergantung pada latar belakang mereka. Penyakit ini bisa dengan cepat menyebar antar wilayah, sehingga diperlukan sistem yang

tepat untuk mengelompokkan data kasus TB di berbagai daerah di provinsi Aceh.

Karena penyebaran TB yang tidak merata serta perbedaan faktor geografis, sosial, dan ekonomi yang memengaruhinya, pemetaan penyebaran TB menjadi tantangan tersendiri (Aem Ismail, 2023). Analisis data spasial kini menjadi salah satu metode yang berkembang dalam beberapa tahun terakhir untuk memahami dinamika penyebaran penyakit dengan lebih menyeluruh. Dalam mengolah data dengan variasi kepadatan dan untuk mengidentifikasi pola penyebaran yang tidak teratur, dibutuhkan metode *clustering* yang lebih canggih. Hal ini disebabkan metode konvensional sering kurang efektif dalam menangani data yang tidak seragam dan mengandung nilai ekstrem atau *outlier*.

Metode DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) dan OPTICS (*Ordering Points To Identify the Clustering Structure*) merupakan dua metode clustering yang dapat digunakan dalam analisis spasial. DBSCAN adalah metode yang berbasis pada kepadatan yang mampu mendeteksi cluster dengan bentuk yang tidak teratur serta mengidentifikasi titik-titik data yang dianggap *noise* atau tidak termasuk dalam cluster mana pun (Ester et al., 1996). Sedangkan OPTICS adalah metode yang lebih fleksibel dan tidak memerlukan penentuan jumlah cluster sebelumnya, memungkinkan deteksi cluster dengan kepadatan yang bervariasi dalam data spasial yang lebih kompleks (Ankerst et al., 1999). Kedua metode ini memiliki keunggulan dalam mengatasi masalah cluster yang tidak terstruktur, tetapi masing-masing memiliki pendekatan yang berbeda dalam mengelompokkan data.

Walaupun metode DBSCAN dan OPTICS telah banyak digunakan dalam berbagai bidang, perbandingan antara kedua metode ini dalam konteks pemetaan penyebaran penyakit TB di Provinsi Aceh masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan untuk membandingkan efektivitas metode DBSCAN dan OPTICS dalam pemetaan penyebaran penyakit tuberkulosis di Provinsi Aceh. Dengan membandingkan kedua metode ini, penelitian ini berusaha untuk menentukan metode yang lebih tepat untuk analisis spasial dalam konteks pencegahan dan penanggulangan TB.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai metode yang lebih akurat dan efisien dalam pemetaan penyebaran penyakit TB di Aceh. Selain itu, penelitian ini juga diharapkan dapat menjadi referensi bagi pengembangan metodologi analisis spasial dalam penelitian kesehatan masyarakat, serta memberikan kontribusi dalam peningkatan strategi pengendalian penyakit TB di Indonesia.

## 1.2. Rumusan Masalah

Rumusan masalah penelitian ini adalah

1. Bagaimana efektivitas metode DBSCAN dalam mengelompokkan wilayah penyebaran penyakit tuberkulosis di Provinsi Aceh.
2. Bagaimana efektivitas metode OPTICS dalam mengelompokkan wilayah penyebaran penyakit tuberkulosis di Provinsi Aceh.
3. Metode clustering manakah yang lebih sesuai untuk digunakan dalam pemetaan penyebaran penyakit tuberkulosis di Provinsi Aceh.
4. Bagaimana menyajikan hasil clustering penyebaran penyakit *tuberculosis* di provinsi Aceh dalam bentuk peta

## 1.3. Tujuan Penelitian

Adapun tujuan penelitian ini adalah:

1. Mengevaluasi kinerja metode DBSCAN dalam proses pengelompokan wilayah penyebaran Tuberkulosis di Provinsi Aceh.
2. Mengevaluasi kinerja metode OPTICS dalam pengelompokan wilayah penyebaran Tuberkulosis di Provinsi Aceh.
3. Melakukan perbandingan hasil *clustering* yang diperoleh dari metode DBSCAN dan OPTICS untuk mengidentifikasi metode yang paling tepat digunakan dalam pemetaan penyebaran Tuberkulosis di Provinsi Aceh.
4. Menyajikan hasil pemetaan penyebaran penyakit tuberkulosis berdasarkan hasil clustering dalam bentuk visualisasi peta yang informatif.

#### **1.4. Batasan Masalah**

Adapun batasan masalah dari penelitian ini adalah:

1. Penelitian ini hanya terbatas pada Provinsi Aceh dengan rentang waktu 2019 sampai 2023, dan tidak mencakup daerah lain di Indonesia atau luar negeri. Analisis dan rekomendasi penelitian ini hanya berlaku untuk wilayah Aceh.
2. Algoritma yang digunakan dalam penelitian ini adalah DBSCAN (*Density-Based Spatial Clustering Of Applications With Noise*) dan OPTICS (*Ordering Points To Identify the Clustering Structure*) tanpa penerapan algoritma tambahan lainnya.
3. Hasil pengelompokan divisualisasikan melalui peta sederhana berdasarkan titik koordinat (latitude dan longitude). Peta hanya berfungsi sebagai media ilustrasi untuk memperjelas hasil analisis, bukan sebagai kajian Sistem Informasi Geografis (SIG) secara komprehensif.

#### **1.5. Manfaat Penelitian**

Adapun manfaat dari penelitian ini adalah:

1. Memperluas kajian ilmiah untuk pemetaan dan analisis kejadian persebaran kasus TB berdasarkan heterogenitas masyarakat.
2. Dapat dilakukan tindakan pencegahan tertentu untuk mencegah penularan wabah terutama pada daerah-daerah yang memiliki banyak kasus aktif.
3. Membantu instansi untuk mengetahui lokasi-lokasi yang memiliki perkembangan pesat kasus TB yang dapat digunakan untuk menentukan kebijakan.

## BAB II

### TINJAUAN KEPUSTAKAAN

#### 2.1 Penelitian Terdahulu

Untuk mencegah plagiarisme, penggunaan referensi atau studi terkait sangat penting dalam penelitian yang akan dilakukan. Studi sebelumnya berperan sebagai acuan, tolak ukur, sekaligus bahan perbandingan. Selain itu, penulis menguraikan temuan-temuan dari penelitian sebelumnya dalam landasan teori untuk mengidentifikasi perbedaan dan memberikan konteks bagi penelitian yang sedang dilakukan.

Tabel 2. 1 Tabel Penelitian Terdahulu

PENELITI (TAHUN)	JUDUL PENELITIAN	HASIL PENELITIAN
Budiman Sujatmiko, (2023)	Analisis Metode Clustering Dengan Algoritma <i>Density – Based Spatial Clustering Of Application With Noise</i> (Dbscan) Untuk Pengelompokan <i>Cyberbullying</i> Di Instagram	Dengan menggunakan algoritma DBSCAN, penelitian berhasil mengelompokkan kalimat cyberbullying di Instagram. Pembobotan nilai pada setiap kata di dalam dokumen berhasil dilakukan menggunakan proses <i>term weighting</i> atau metode TF/IDF. Dalam implementasi DBSCAN, penentuan nilai Epsilon sebesar 0,5 dan minpoint 3

		<p>menghasilkan kluster dengan nilai -1.</p> <p>Pengujian algoritma DBSCAN menggunakan Davies-Bouldin Index memberikan hasil dengan nilai 1.</p>
<p>Dedy Armiady, (2022)</p>	<p>Analisis Metode DBSCAN (<i>Density-Based Spatial Clustering of Application with Noise</i>) dalam Mendeteksi Data <i>Outlier</i></p>	<p>Penelitian ini menyimpulkan bahwa DBSCAN efektif mendeteksi <i>outlier</i> melalui konsep <i>density-reachable</i>. Satu data dalam dataset dengan nilai <math>x=0,370007</math> dan <math>y=0,410475</math> terdeteksi sebagai <i>outlier</i> di semua model, ditandai dengan warna abu-abu pada <i>scatter plot</i>. Semakin besar nilai Epsilon, jumlah kluster berkurang karena jangkauan <i>core point</i> dan <i>border point</i> meningkat. Meski begitu, DBSCAN dinilai kurang optimal karena beberapa kluster memiliki jarak</p>

		<p>antar titik yang terlalu jauh. Penulis merekomendasikan metode dengan nilai Epsilon dinamis untuk hasil kluster yang lebih akurat.</p>
<p>(Setiyawan &amp; Barkah, 2025)</p>	<p>Comparative Analysis of DBSCAN, OPTICS, and Agglomerative Clustering Methods for Identifying Disease Distribution Patterns in Banjarnegara Community Health Centers Dillyana</p>	<p>Penelitian ini membandingkan metode DBSCAN, OPTICS, dan <i>Agglomerative Clustering</i> dalam menganalisis pola penyakit di Puskesmas Banjarnegara. Hasil menunjukkan bahwa <i>Agglomerative Clustering</i> paling efektif karena menghasilkan kluster yang stabil dan mudah diinterpretasi. DBSCAN menghasilkan skor evaluasi tertinggi namun sensitif terhadap parameter dan banyak menghasilkan <i>noise</i>, sedangkan OPTICS kurang stabil dengan</p>

		<p>banyak <i>outlier</i>. Secara keseluruhan, metode hierarkis lebih cocok untuk mendukung kebijakan kesehatan berbasis data.</p> <p>Tanya ChatGPT</p>
Dwina Pri Indini, et al., (2022)	<p>Implementasi Algoritma DbSCAN Untuk Clustering Seleksi Penentuan Mahasiswa Yang Berhak Menerima Beasiswa Yayasan</p>	<p>Penelitian ini menyimpulkan bahwa algoritma DBSCAN efektif membentuk kluster berdasarkan kerapatan, menghasilkan dua kluster. Mahasiswa yang layak menerima beasiswa berada di Kluster I, yaitu alternatif D, F, G, H, I, dan J. Hasil ini dapat digunakan sebagai rekomendasi untuk memudahkan dan mempercepat proses seleksi beasiswa.</p>

(Hastuti et al., 2024)	Penerapan Metode OPTICS dan ST-DBSCAN untuk Klasterisasi Data Kesehatan	Penelitian ini menunjukkan bahwa metode OPTICS dan ST-DBSCAN mampu mengelompokkan kecamatan berdasarkan distribusi tenaga kesehatan. Namun, OPTICS memberikan hasil kluster lebih optimal, konsisten, dan informatif dengan nilai silhouette lebih tinggi, sehingga efektif menangani data kesehatan yang mengandung derau dan outlier.
Dia Ayu Puspitasari et al., (2023)	Penerapan Algoritma Density Based Spastial Clustering Algorithm With <i>Noise</i> Untuk Pengelompokan Penyakit Y Pasien.	Penelitian ini menyimpulkan bahwa metode DBSCAN berhasil diimplementasikan pada dataset rekam medis, menghasilkan nilai Davies-Bouldin Index sebesar 1.42. Pengelompokan penyakit

		<p>pasien menghasilkan dua kluster menggunakan Python di Jupyter, sementara perhitungan manual menunjukkan tiga iterasi dengan core point di 4, 38, dan 45.</p>
<p>(Setiyawan &amp; Barkah, 2025)</p>	<p>Comparative Analysis of DBSCAN, OPTICS, and Agglomerative Clustering Methods for Identifying Disease Distribution Patterns in Banjarnegara Community Health Centers</p>	<p>Penelitian ini menyimpulkan bahwa Agglomerative Clustering paling efektif mengelompokkan Puskesmas berdasarkan distribusi penyakit karena menghasilkan kluster stabil dan mudah diinterpretasikan. DBSCAN dan OPTICS berguna mendeteksi outlier, namun sensitif terhadap parameter dan karakteristik data.</p>

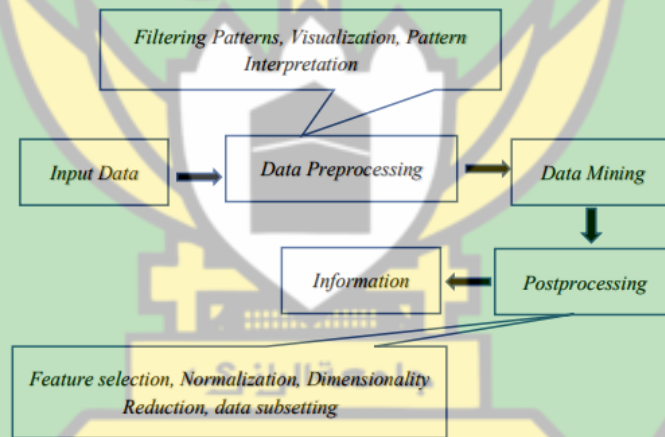
## 2.2 Tuberkulosis (TB)

Tuberkulosis (TB) Infeksi yang disebabkan oleh bakteri *Mycobacterium tuberculosis*, yang biasanya menyerang paru-paru, menyebabkan gangguan pernapasan (Natarajan et al., 2020; Panic et al., 2020; Widiyanto, 2017; Iswahyuni, 2022). Bakteri TB dapat menyerang kelenjar getah bening, pankreas, ginjal, tulang, dan juga organ

tubuh lainnya saat menjadi lebih parah. Gejala yang ditunjukkan oleh TB hampir mirip dengan gejala penyakit pernapasan lainnya. Mereka juga mengalami penurunan nafsu makan dan berat badan yang buruk, demam berdarah, dan berkeringat pada malam hari tanpa alasan selama lebih dari tiga minggu. Penularan TB paling mungkin terjadi pada orang yang memiliki kekebalan rendah, seperti penderita HIV atau AIDS, bayi atau anak-anak, orang dengan penyakit berat dan masalah gizi, seperti diabetes melitus, gagal ginjal, penyakit jantung, dan orang lanjut usia (Di & Tunggul, 2023).

### 2.3 Data Mining

Data mining adalah proses eksplorasi dan ekstraksi pengetahuan dari sejumlah besar data, baik itu dari database maupun repositori lainnya. Tujuan utama dari proses ini adalah mengungkap pengetahuan baru yang tersembunyi di dalam data tersebut. Proses ini juga dikenal sebagai upaya untuk menemukan informasi yang sebelumnya tidak teridentifikasi secara manual dari kumpulan data.



Gambar 2. 1 Diagram Proses Data Mining

Berikut merupakan penjelasan dari tahapan data mining oleh (Tan et al., 2020) pada Gambar 2. 1, sebagai berikut:

#### 1. Data Selection

Pemilihan data adalah penentuan jenis data, sumber data, dan instrumen data yang tepat yang dapat digunakan. Hasil pemilihan tersebut yang akan

digunakan untuk proses data mining disimpan dalam berkas terpisah dari basis data operasional.

## 2. *Cleaning*

Proses pembersihan data meliputi beberapa langkah, seperti menghapus data yang duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pengetikan atau typo.

## 3. *Transformation*

Data mining membutuhkan perubahan format data agar dapat diproses. Beberapa metode standar, seperti clustering dan analisis asosiasi, hanya dapat menerima data kategorik, sehingga data yang berupa angka numerik terus-menerus harus dibagi menjadi beberapa interval.

## 4. Proses Data Mining

Proses menemukan pola atau informasi yang menarik dalam data yang dipilih dengan menggunakan metode atau teknik tertentu. Teknik clustering digunakan dalam proses data mining penelitian ini.

## 5. *Pattern Evaluation*

Tahap ini menentukan apakah informasi atau pola yang ditemukan bertentangan dengan temuan atau teori sebelumnya.

Muncul dari berbagai bidang keilmuan, termasuk kecerdasan buatan, basis data, statistik, dan pemrosesan gambar digital. Data mining adalah topik yang relatif muda dalam ilmu komputer. Penerapan data mining semakin meluas sebagai hasil dari integrasi berbagai bidang ini. (Subariah & Zein, 2023). Data mining terbagi ke dalam berbagai kategori sesuai dengan tugas yang dapat diselesaikan.

1. *Description* (Deskripsi)

Biasanya para peneliti hanya mencari cara untuk mengidentifikasi cara-cara untuk mengkarakterisasi tren dan pola dalam data. Dalam pemilihan presiden, misalnya, seorang peneliti jajak pendapat dapat mengalami kesulitan menemukan data atau fakta mengenai siapa yang dipandang kurang profesional (Armiady, 2022).

2. *Estimation* (Estimasi)

Estimasi hampir sama dengan klasifikasi, dalam estimasi variabel target biasanya berbentuk numerik dan bukan kategorikal. Ketika melakukan estimasi, sebuah model dibuat dengan menggunakan semua informasi yang tersedia, termasuk nilai dan variabel target yang diprediksi. Memperkirakan indeks prestasi kumulatif (IPK) mahasiswa pascasarjana berdasarkan IPK sarjana mereka adalah contoh bagaimana estimasi diterapkan (Armiady, 2022).

3. *Prediction* (Prediksi)

Prediksi memiliki kesamaan dengan proses estimasi dan klasifikasi. Misalnya, mengantisipasi curah hujan selama tujuh hari ke depan atau harga emas selama tiga bulan ke depan. Prediksi sering kali menggunakan teknik dan prosedur yang sama dengan yang digunakan dalam estimasi dan kategorisasi. Diharapkan bahwa teknik penggalian data akan menghasilkan informasi prediktif mengenai status desa dengan menggunakan data Indeks Pembangunan Desa (Armiady, 2022).

4. *Classification* (Klasifikasi)

Dalam data mining, klasifikasi adalah metode penting untuk mengelompokkan data ke dalam beberapa kategori. Metode ini memanfaatkan data pelatihan yang telah memiliki label kelas yang ditetapkan untuk setiap atribut. Label kelas adalah variabel target yang digunakan dalam proses klasifikasi. Selanjutnya, sekumpulan data yang berisi informasi tentang variabel target dan sekumpulan variabel input atau prediktor dievaluasi oleh model klasifikasi (Armiady, 2022).

## 5. *Clustering* (Pengklasteran)

Clustering adalah praktik atau cara mengatur suatu informasi atau item sesuai dengan karakteristik yang sama.. Sebuah kelompok (cluster) terdiri dari data yang memiliki karakteristik serupa satu sama lain, namun berbeda dari data yang ada di kelompok lain. Contoh penerapannya adalah dalam mencari kelompok konsumen untuk penargetan dalam pemasaran produk, terutama bagi perusahaan dengan anggaran pemasaran terbatas (Armiady, 2022).

## 6. *Association* (Asosiasi)

Dalam penggalian data, asosiasi adalah proses mengidentifikasi hubungan antara fitur-fitur yang muncul bersamaan. Analisis konsumen perusahaan telekomunikasi yang cenderung menerima tawaran peningkatan layanan adalah ilustrasi bagaimana metode ini digunakan dalam bisnis dan penelitian (Sujatmiko, 2023).

Dari 6 pembagian tugas data mining, peneliti menggunakan clustering untuk penelitian ini. peneliti menggunakan clustering untuk pengelompokan klaster suatu daerah di provinsi Aceh berdasarkan jumlah kasus TB menggunakan metode DBSCAN

## 2.4 Clustering

Salah satu teknik data mining yang digunakan untuk menemukan kelompok item dengan kategori yang mirip dalam volume data yang sangat besar adalah clustering. Tujuan utama metode clustering adalah untuk mengatur berbagai objek atau data ke dalam kelompok-kelompok, yang masing-masing terdiri dari data yang sangat mirip satu sama lain. Sebuah klaster terdiri dari sekelompok data yang berbeda dari data di klaster lainnya, namun memiliki sifat yang sama. Karena pengelompokan didasarkan pada kemiripan item, metode ini bersifat *unsupervised learning* karena pengelompokannya berdasarkan kesamaan antar objek (Asyuti & Setyawan, 2023).

## 2.5 Algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*)

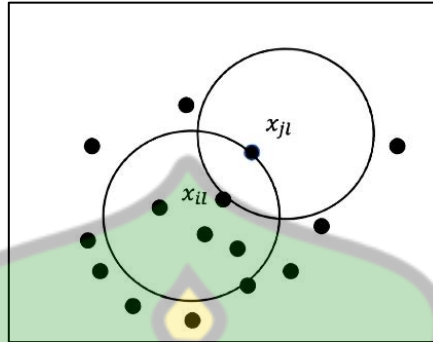
Algoritma DBSCAN (*Density-based spatial clustering of applications with noise*) adalah algoritma clustering berbasis kepadatan, atau clustering berbasis kepadatan. Algoritma ini menganggap cluster didefinisikan sebagai wilayah dengan objek padat dibagi dengan area yang kurang padat. DBSCAN mendefinisikan cluster sebagai kumpulan terbesar dari titik-titik yang saling terhubung secara padat. Setiap item yang tidak termasuk dalam cluster dianggap sebagai *noise*. Titik-titik yang berada dalam radius tertentu dan memiliki jumlah titik minimum (*MinPts*) digunakan untuk pengelompokan (Okfalisa et al., 2021).

Berikut beberapa bagian dari metode DBSCAN antara lain:

1. *Core point*: Titik pusat dalam suatu klaster yang didasarkan pada densitas, di mana sejumlah titik data harus berada dalam radius tertentu (*Epsilon*). Nilai *minPts* (jumlah minimum titik dalam klaster) ditentukan oleh pengguna untuk menentukan apakah titik tersebut memiliki cukup "tetangga" dalam radius *Eps* agar dapat dianggap sebagai titik inti dari klaster.
2. *Border point*: Titik yang berada di dalam jangkauan radius *Eps* dari titik inti (*core point*), tetapi tidak memiliki cukup tetangga untuk dianggap sebagai titik inti sendiri. Titik ini merupakan bagian dari klaster, tetapi berada di tepi atau batas dari klaster tersebut.
3. *Noise point*: Titik yang tidak dapat dijangkau oleh titik inti (*core point*) dan tidak termasuk dalam batas klaster (*border point*). Titik ini dianggap sebagai outlier atau data yang tidak terkelompok dalam klaster mana pun.

$$Noise = \{ x \in X \mid \forall i : x \in C_i \} \quad (1)$$

4. *Directly density-reachable*: *Directly density-reachable* dalam DBSCAN mengacu pada situasi di mana titik A dapat terhubung langsung ke titik B

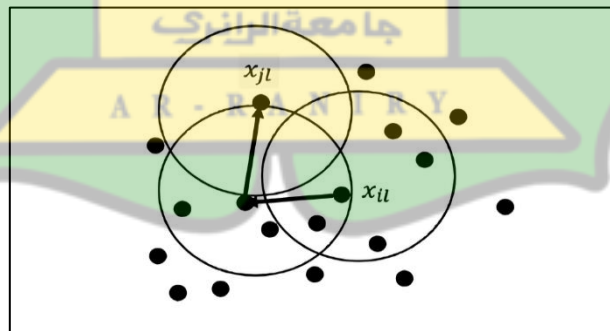


Gambar 2. 2 Ilustrasi Directly density-reachable

jika titik B merupakan *core point* (memiliki cukup banyak tetangga dalam radius tertentu) dan titik A berada dalam jarak radius tersebut. Dengan kata lain, kedua titik ini cukup dekat satu sama lain, sehingga dapat digolongkan dalam kluster yang sama berdasarkan kepadatan data di sekitarnya

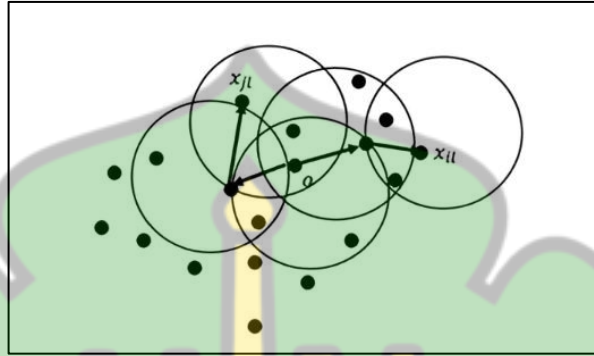
$$x \in N_{Eps}(y) \wedge |N_{Eps}(y)| \geq MinPts \quad (2)$$

5. *Density-reachable*: kondisi *density-reachable* terjadi ketika dua titik saling terhubung melalui titik data lain yang memiliki sifat *directly density-reachable* terhadap kedua titik tersebut. Kondisi ini dapat digambarkan pada Gambar.



Gambar 2. 3 Ilustrasi Density-reachable

6. *Density-connected*: Suatu titik dikatakan saling terhubung berdasarkan densitas jika titik tersebut dapat terhubung satu sama lain melalui satu atau lebih titik lain. Algoritma DBSCAN memerlukan dua parameter input, yaitu epsilon (*Eps*) dan minimum points (*minPts*).



Gambar 2. 4 Ilustrasi *Density-connected*

Titik-titik yang berada dalam radius *Eps* didefinisikan sebagai titik sekitar di definisikan sebagai:

$$N_{Eps}(x) = \{y \in D \mid dist(x, y) \leq Eps\} \quad (3)$$

$N_{Eps}(x)$  = titik sekitar dari  $x$  dalam radius *Eps*

$D$  = gugus data

$dist(x, y)$  = jarak *Euclidean* dari objek  $x$  dan  $y$

*Eps* = radius atau ambang batas.

Untuk melakukan pengelompokan data, langkah pertama yang harus dilakukan adalah menetapkan nilai *minPts* dan *epsilon*. Algoritma DBSCAN biasanya memulai prosesnya dengan memilih titik secara acak sebagai titik awal.

Langkah-langkah metode DBSCAN:

1. Tentukan titik awal atau  $p$  secara acak.
2. Hitung semua jarak titik menggunakan persamaan 4

$$E(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (4)$$

3. Untuk menentukan nilai *epsilon* dan *minPts*, diperlukan beberapa kali percobaan. Nilai *epsilon* dan *minPts* yang optimal adalah yang memberikan *silhouette index* tertinggi dan menghasilkan jumlah kluster yang lebih banyak. Proses ini melibatkan pengujian berulang untuk menemukan kombinasi terbaik dari kedua parameter tersebut.

Keuntungan dari algoritma ini ialah (Shella Novi, 2020):

1. DBSCAN tidak membutuhkan informasi mengenai jumlah kelompok yang akan dibentuk sebelumnya.
2. DBSCAN hanya memerlukan dua parameter yang umumnya tidak terlalu dipengaruhi oleh basis data yang digunakan..

## 2.6 Algoritma OPTICS (Ordering Points To Identify the Clustering Structure)

OPTICS (Ordering Points To Identify the Clustering Structure) adalah metode clustering yang lebih fleksibel dibandingkan dengan DBSCAN. Algoritma OPTICS dikembangkan sebagai penyempurnaan dari algoritma DBSCAN. Untuk meminimalkan ketergantungan terhadap parameter yang digunakan dalam DBSCAN, algoritma ini menentukan nilai *Eps* untuk setiap titik berdasarkan nilai *MinPts* yang telah ditetapkan, lalu memetakannya ke dalam grafik. Dengan cara ini, pengguna dapat mengidentifikasi kluster sesuai dengan nilai *Eps* yang diinginkan (TURGUT et al., 2022).

Seperti DBSCAN, OPTICS juga berbasis pada kepadatan, namun tidak memerlukan penentuan jumlah cluster yang tetap sebelumnya. OPTICS menghasilkan urutan titik berdasarkan kedekatan dan kepadatan data, yang memungkinkan untuk menggambarkan struktur clustering yang lebih kompleks (Aryo Wibisono Putra, Zulham Ariyandi, Revi Diwana, Supriyatna, 2024).

Kelebihan utama OPTICS adalah kemampuannya untuk mengidentifikasi cluster dengan kepadatan yang bervariasi dan tidak teratur dalam data yang lebih kompleks. Selain itu, OPTICS dapat memberikan informasi tentang kualitas cluster, sehingga lebih mudah untuk menyesuaikan analisis dengan karakteristik data yang ada (Ester et al., 1996). Namun, metode ini juga memerlukan pemahaman yang lebih mendalam dalam mengatur parameter dan interpretasi hasil clustering. (Mustikasari & Salman, 2023), beberapa istilah dalam metode OPTICS antara lain :

1. Core Distance

Core distance adalah jarak minimum yang diperlukan agar sebuah titik dapat menghubungkan sejumlah titik lainnya dalam radius yang telah ditentukan. Titik yang memiliki core distance kecil dan cukup banyak tetangga dalam jarak tersebut disebut core point. Fungsi Core distance digunakan untuk menentukan apakah suatu titik dapat menjadi pusat klaster atau tidak.

2. Reachability Distance

Reachability distance adalah jarak antara dua titik yang menunjukkan seberapa mudah kedua titik tersebut saling terhubung, dengan mempertimbangkan jarak antar titik dan core distance mereka. Fungsi Reachability distance mengukur kedekatan dua titik dalam pengklasteran, semakin kecil nilai ini, semakin besar kemungkinan kedua titik tersebut berada dalam klaster yang sama.

### 3. Nilai $\xi$ (xi)

Nilai ini digunakan sebagai ambang batas dalam proses ekstraksi klaster dari *reachability plot*, untuk mendeteksi penurunan signifikan dalam jarak antar titik yang menandai batas antar klaster. Pemilihan nilai 0.05 mengacu pada standar yang umum digunakan dalam literatur untuk menjaga keseimbangan antara sensitivitas dan kestabilan hasil klasterisasi.

## 2.7 Pengujian Indeks Silhouette

Pengujian indeks Silhouette digunakan untuk mengevaluasi akurasi clustering. Analisis metode ini dilakukan dengan mengamati nilai indeks Silhouette (s), yang berkisar antara -1 hingga 1, tergantung pada seberapa baik objek dikelompokkan. Nilai Silhouette yang mendekati 1 menunjukkan bahwa objek berada dalam cluster yang sesuai. Sebaliknya, jika nilai mendekati 0, berarti objek tidak memiliki kepastian mengenai kelompoknya. Sedangkan nilai yang mendekati -1 menandakan bahwa objek kemungkinan berada di cluster yang salah (Rousseeuw, 1987). Untuk menghitung Indeks Silhouette dapat digunakan dengan persamaan 5.

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i) \quad (5)$$

Dimana:

$sil(k)$  : nilai silhouette semua

$|k|$  : banyaknya cluster k

$sil c_i$  : rata-rata nilai silhouette

## **BAB III**

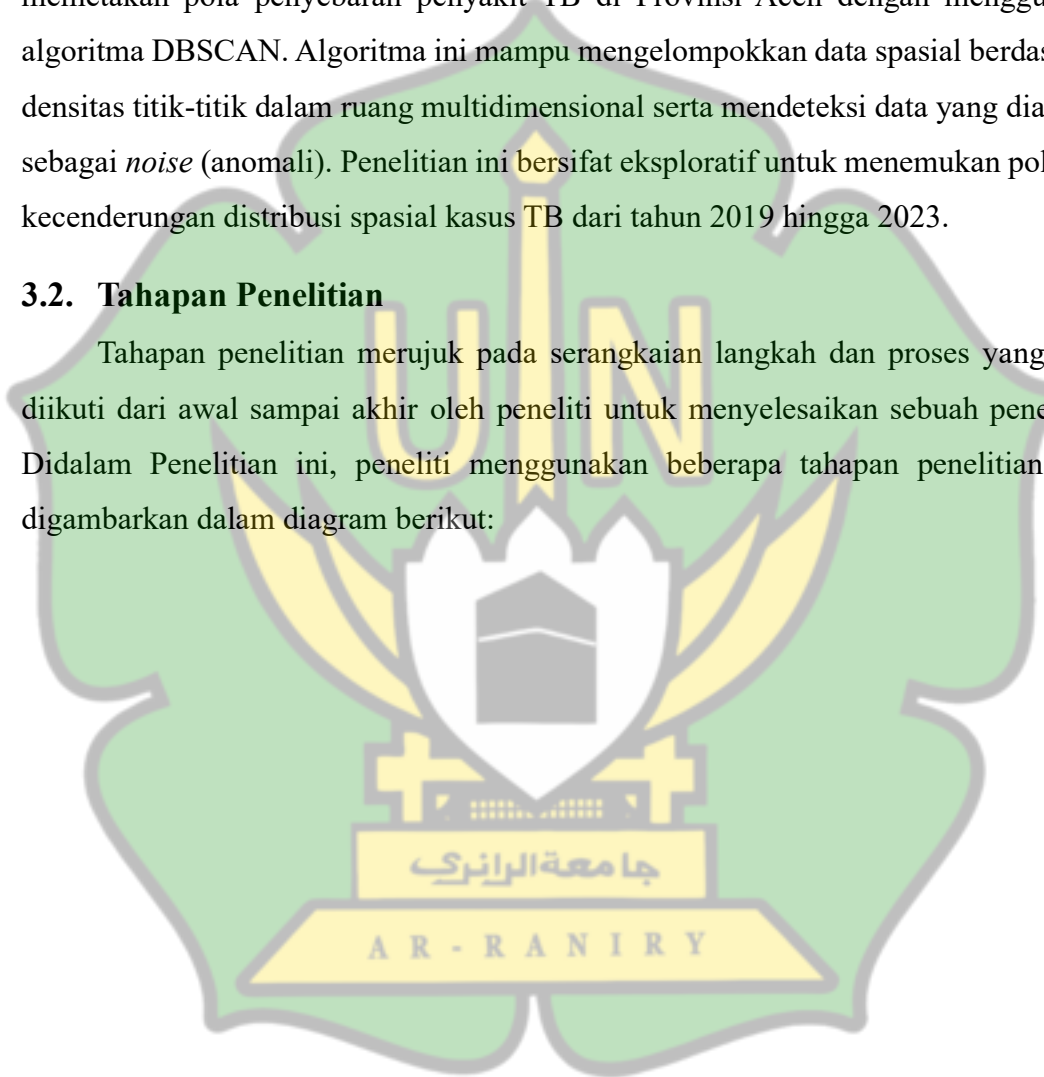
### **METODE PENELITIAN**

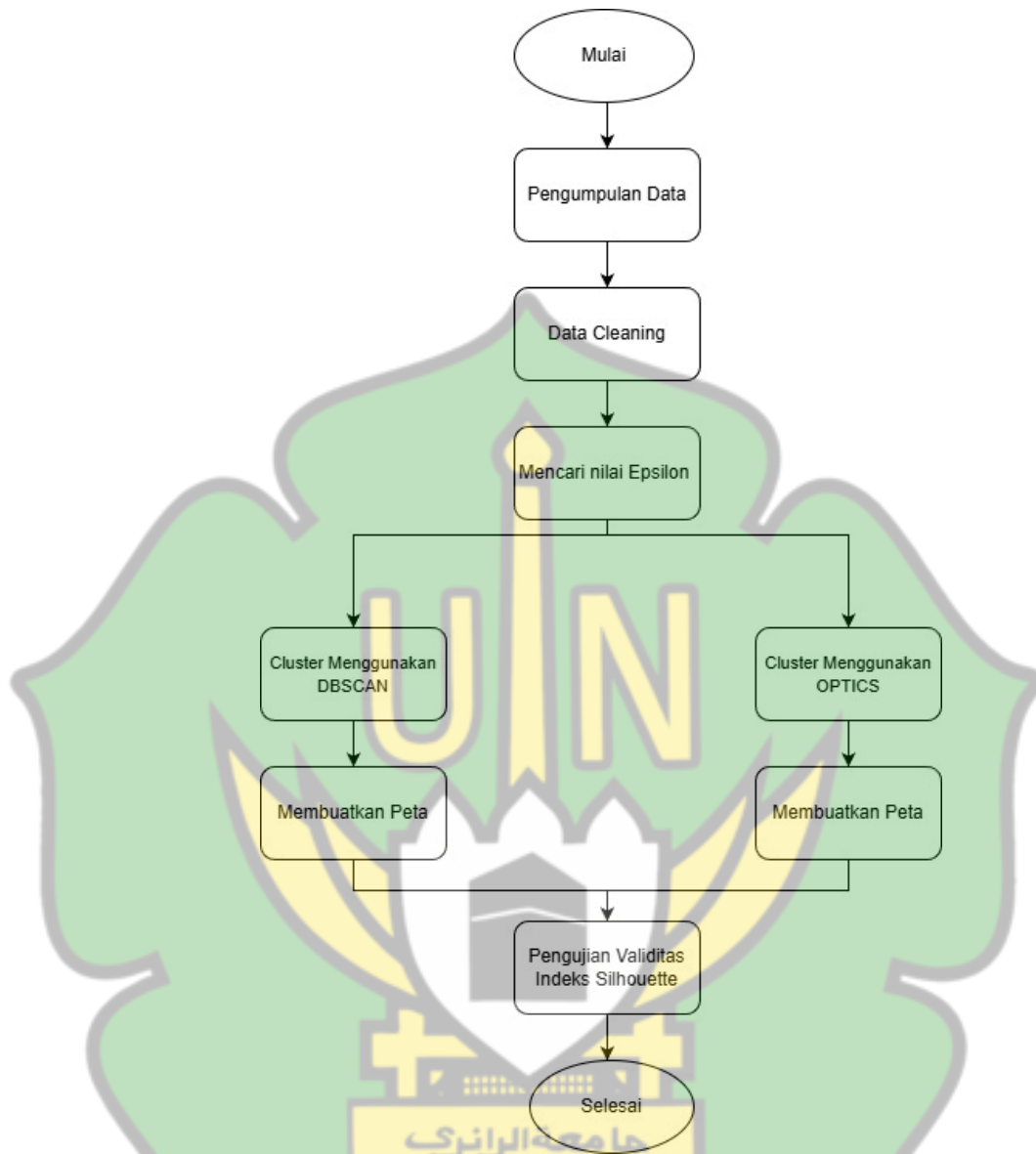
#### **3.1. Jenis Penelitian**

Penelitian ini merupakan penelitian deskriptif kuantitatif yang bertujuan untuk memetakan pola penyebaran penyakit TB di Provinsi Aceh dengan menggunakan algoritma DBSCAN. Algoritma ini mampu mengelompokkan data spasial berdasarkan densitas titik-titik dalam ruang multidimensional serta mendeteksi data yang dianggap sebagai *noise* (anomali). Penelitian ini bersifat eksploratif untuk menemukan pola atau kecenderungan distribusi spasial kasus TB dari tahun 2019 hingga 2023.

#### **3.2. Tahapan Penelitian**

Tahapan penelitian merujuk pada serangkaian langkah dan proses yang perlu diikuti dari awal sampai akhir oleh peneliti untuk menyelesaikan sebuah penelitian. Didalam Penelitian ini, peneliti menggunakan beberapa tahapan penelitian yang digambarkan dalam diagram berikut:





Gambar 3. 1 Diagram Alir Penelitian

### 3.2.1. Pengumpulan Data

Data mengenai penyebaran penyakit TB dikumpulkan dari berbagai sumber yang relevan, seperti data dinas kesehatan, laporan epidemiologi, atau survei lapangan. Data ini meliputi informasi terkait lokasi pasien, jumlah kasus, serta tahun data diperoleh.

Tabel 3. 1 Data TB Aceh 2019-2023

Kabupaten	2019	2020	2021	2022	2023
Aceh Barat	311	259	184	290	291
Aceh Barat Daya	202	232	247	262	315
Aceh Besar	408	347	355	437	439
Aceh Jaya	212	155	120	143	148
Aceh Selatan	374	281	415	747	793
Aceh Singkil	208	95	67	196	197
Aceh Tamiang	426	259	447	451	453
Aceh Tengah	180	189	147	266	272
Aceh Tenggara	129	58	160	345	365
Aceh Timur	567	578	648	813	821
Aceh Utara	901	762	671	1.016	1.024
Banda Aceh	997	379	604	1.108	1.156
Bener Meriah	67	100	62	104	115
Bireuen	848	740	729	876	881
Gayo Lues	199	105	149	226	235
Langsa	410	249	304	485	495
Lhokseumawe	598	293	587	853	882
Nagan Raya	284	123	180	373	384
Pidie	799	587	718	1.144	1.151
Pidie Jaya	199	65	85	161	168
Sabang	30	18	9	28	31
Simeulue	124	149	80	201	213
Subulussalam	174	72	202	371	415

### 3.2.2. Data Cleaning

Data yang telah terkumpul dibersihkan untuk memastikan kualitasnya. Data yang tidak lengkap, tidak relevan, atau memiliki kesalahan diidentifikasi dan diperbaiki. Misalnya, titik data dengan koordinat yang salah atau duplikasi

kasus akan dihapus atau diperbaiki. Ini penting untuk memastikan bahwa analisis yang dilakukan di kemudian hari didasarkan pada data yang akurat.

### 3.2.3. Mencari Nilai Epsilon

Pada tahap ini, nilai *epsilon* ( $\epsilon$ ) ditentukan sebagai salah satu parameter utama dalam algoritma DBSCAN. Nilai *epsilon* adalah jarak maksimum yang digunakan untuk menentukan apakah dua titik data dapat dianggap berada dalam satu cluster. Penentuan nilai *epsilon* yang tepat penting untuk menghasilkan kluster yang sesuai dengan distribusi geografis penyebaran TB di Aceh.

### 3.2.4. Clustering dengan Metode DBSCAN

Algoritma DBSCAN diterapkan pada dataset penyebaran TB untuk mengidentifikasi area dengan konsentrasi tinggi kasus TB. DBSCAN merupakan metode yang cocok untuk mendeteksi kluster berdasarkan kepadatan dan mampu menangani outlier atau titik data yang tidak sesuai dengan pola umum. Dalam konteks penelitian ini, DBSCAN akan membantu mengidentifikasi wilayah dengan tingkat penyebaran TB yang tinggi.

### 3.2.5. Clustering dengan Metode OPTICS

Metode OPTICS (*Ordering Points To Identify the Clustering Structure*) digunakan untuk mengelompokkan data sebaran kasus TB berdasarkan variasi tingkat kepadatan antar wilayah. Berbeda dari DBSCAN, OPTICS tidak memerlukan penentuan jarak ambang (*epsilon*) secara eksplisit, sehingga lebih adaptif dalam mendeteksi kluster dengan kepadatan yang tidak seragam. Dalam penelitian ini, OPTICS dimanfaatkan untuk mengenali area-area dengan karakteristik penyebaran TB yang mirip serta mengidentifikasi titik-titik yang menyimpang dari pola mayoritas. Dengan memanfaatkan hasil *reachability plot*, metode ini membantu menggambarkan struktur sebaran penyakit TB secara lebih rinci di masing-masing Puskesmas.

### **3.2.6. Pengujian Validitas Indeks Silhouette**

Setelah proses klusterisasi selesai, validitas hasil clustering dievaluasi menggunakan indeks Silhouette. Indeks ini mengukur seberapa baik klaster yang terbentuk, apakah klaster-klaster tersebut saling terpisah dengan baik dan seberapa dekat anggota klaster berada dengan pusat klaster mereka. Nilai indeks Silhouette akan memberikan gambaran tentang seberapa akurat clustering dalam memetakan penyebaran TB di Aceh.

### **3.3. Data Dan Variabel Penelitian**

Penelitian ini menemukan dan mengkaji pola pengelompokan kasus tuberkulosis di Provinsi Aceh dengan menggunakan pendekatan kuantitatif dan teknik data mining. Metode clustering dengan algoritma DBSCAN dipilih untuk melakukan pengelompokan data berdasarkan karakteristik yang ada. Penelitian ini menggunakan data sekunder dari E-Profil Kesehatan Aceh dan Global Tuberkulosis Report 2023. Variabel yang digunakan dalam penelitian ini terdiri atas variabel independen dan variabel dependen yang merupakan data numerik dengan skala pengukuran rasio. Terdapat beberapa indikator dengan jumlah seluruh variabel independen sebanyak 5 variabel. Rincian variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3.2.

Tabel 3. 2 Variabel Penelitian

Simbol	Definisi Variabel	Satuan
$x_1$	Latitude	Derajat (°)
$x_2$	Longitude	Derajat (°)
$y_1$	Jumlah kasus TB 2021	Jiwa
$y_2$	Jumlah kasus TB 2022	Jiwa
$y_3$	Jumlah kasus TB 2023	Jiwa
$y_4$	Jumlah kasus TB 2020	Jiwa
$y_5$	Jumlah kasus TB 2019	Jiwa

### 3.4. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Variabel Bebas: Koordinat geografis (longitude dan latitude) dari setiap kasus TB.
2. Variabel Terikat: Pola atau kluster penyebaran kasus TB di Provinsi Aceh.

### 3.5. Data dan Sumber Data

Data yang digunakan dalam penelitian ini meliputi data sekunder yang berasal dari:

1. Data Kasus TB: Data kasus Tuberkulosis di Provinsi Aceh pada periode 2019–2023, diperoleh dari dinas kesehatan provinsi Aceh, pusat data dan informasi kesehatan (Pusdatin), serta instansi terkait yang menyediakan data kasus TB.
2. Data Geospasial: Data koordinat geografis (latitude dan longitude) lokasi kasus TB yang digunakan dalam analisis spasial. Data ini didapatkan dari data lokasi fasilitas kesehatan, seperti rumah sakit dan puskesmas yang menangani kasus TB, atau dari sumber lain yang relevan.

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 ANALISIS DEKRIPTIF

Analisis deskriptif adalah metode statistik yang digunakan untuk memberikan gambaran umum tentang data yang akan dianalisis.

Tabel 4. 1 Tabel Deskriptif

	Maks	Min	Mean	Total	Standard Deviation
TB 2019	997	30	375.957	8647	280.673
TB 2020	762	18	265	6095	214.3481
TB 2021	729	9	311.739	7170	238.9103
TB 2022	1144	28	473.739	10896	340.9552
TB 2023	1156	31	488.87	11244	345.5764

Berdasarkan hasil analisis pada Tabel 4.1 Pada tahun 2019, jumlah kasus TB tertinggi mencapai 997 kasus dan terendah sebesar 30 kasus. Nilai rata-rata kasus sebesar 375,957 dan total keseluruhan kasus sebanyak 8.647. Simpangan baku sebesar 280,673 mengindikasikan bahwa terdapat perbedaan yang cukup besar dalam distribusi jumlah kasus antar wilayah.

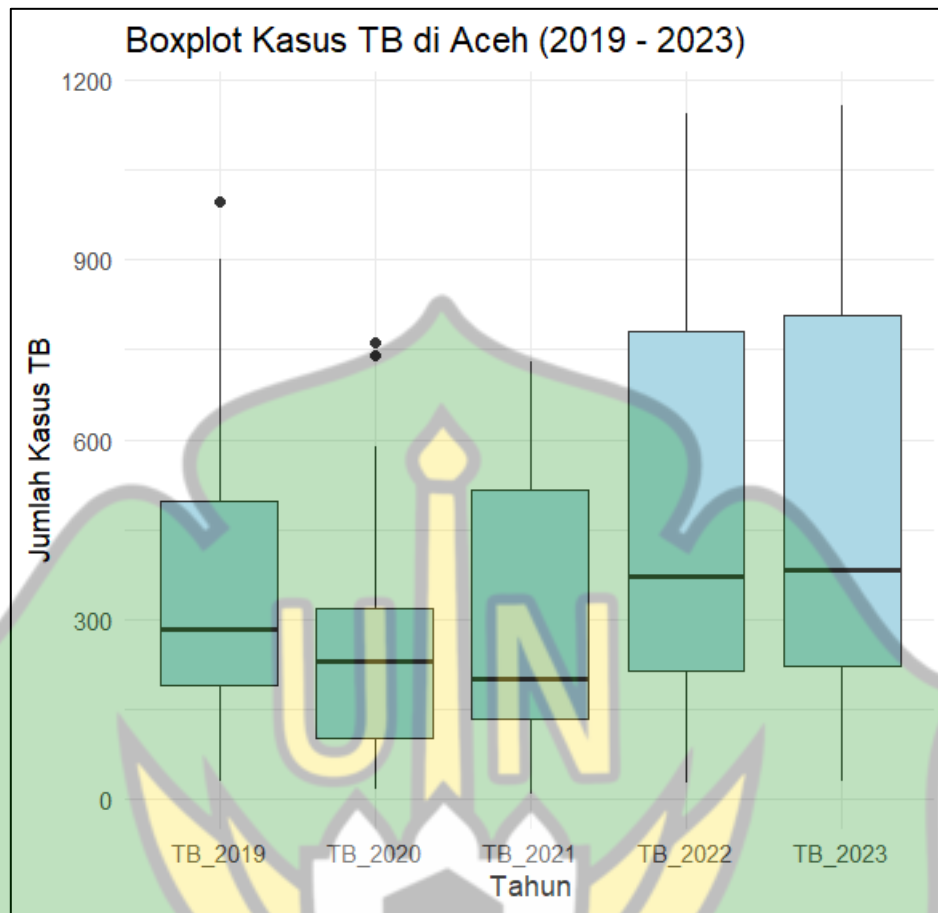
Memasuki tahun 2020, terjadi penurunan tajam baik dari sisi rata-rata maupun total kasus, yang hanya mencapai 6.095 dengan rata-rata 265 kasus. Rentang nilai berkisar antara 18 hingga 762 kasus, dan simpangan baku menurun menjadi 214,3481. Penurunan ini dapat dikaitkan dengan dampak pandemi COVID-19, yang menyebabkan perhatian layanan kesehatan terfokus pada virus corona, sehingga pelaporan dan penanganan TB menjadi terbatas.

Pada tahun 2021, kasus TB kembali mengalami peningkatan. Nilai minimum berada pada angka 9 kasus dan maksimum 729 kasus, dengan rata-rata sebesar 311,739 dan total kasus sebanyak 7.170. Meningkatnya simpangan baku menjadi 238,9103 menunjukkan bahwa ketimpangan antar daerah mulai kembali terlihat meskipun belum terlalu ekstrem.

Tren peningkatan lebih signifikan terjadi pada tahun 2022, dengan jumlah kasus tertinggi mencapai 1.144 dan terendah 28 kasus. Rata-rata kasus naik menjadi 473,739 dan total kasus mencapai 10.896. Simpangan baku sebesar 340,9552 mencerminkan adanya ketidakseimbangan distribusi kasus yang cukup besar di antara wilayah yang diamati.

Puncaknya terjadi pada tahun 2023, di mana jumlah kasus maksimum tercatat sebesar 1.156 dan minimum sebesar 31 kasus. Rata-rata kasus TB meningkat menjadi 488,87 dengan total keseluruhan mencapai 11.244 kasus. Nilai simpangan baku tertinggi selama lima tahun terakhir, yakni 345,5764, menunjukkan adanya ketimpangan yang semakin besar antar daerah dalam jumlah kasus TB.

Secara umum, terdapat pola peningkatan jumlah kasus TB pasca-2020 hingga 2023. Meskipun jumlah total dan rata-rata kasus terus meningkat, nilai simpangan baku yang cukup besar setiap tahunnya menunjukkan bahwa penyebaran kasus tidak merata. Hal ini menandakan perlunya strategi intervensi yang disesuaikan dengan kondisi masing-masing daerah agar pengendalian TB dapat lebih efektif dan merata.



Gambar 4. 1 Boxplot kasus TB 2019 - 2023

Secara keseluruhan, jumlah kasus TB (Tuberkulosis) di Aceh mengalami kenaikan yang signifikan antara tahun 2019 hingga 2023, yang tercermin dari peningkatan median setiap tahunnya, terutama pada 2022 dan 2023. Pada kedua tahun tersebut, terdapat peningkatan yang jelas, dengan rentang data yang lebih lebar dan median yang lebih tinggi, yang menunjukkan adanya lonjakan dan variasi jumlah kasus yang lebih besar dibandingkan dengan tahun-tahun sebelumnya. Hal ini juga disertai dengan peningkatan variabilitas, terlihat dari whiskers yang lebih panjang dan adanya banyak nilai pencilan (outliers), terutama di 2022 dan 2023. Keberadaan outliers ini menandakan adanya lonjakan luar biasa, yang mungkin disebabkan oleh faktor-faktor eksternal seperti perubahan kebijakan kesehatan atau krisis tertentu.

Sebaliknya, periode 2019 hingga 2021 menunjukkan stabilitas dalam jumlah kasus, dengan boxplot pada tahun-tahun tersebut menunjukkan distribusi yang lebih sempit dan median yang lebih rendah, menandakan kontrol yang lebih baik terhadap angka kasus. Peningkatan yang terjadi pada tahun 2022 dan 2023 memerlukan penyelidikan lebih mendalam untuk mengetahui penyebab lonjakan ini, termasuk faktor kebijakan atau aspek sosial-ekonomi yang memengaruhi penyebaran. Hal ini juga menunjukkan pentingnya melakukan tindakan pengendalian yang lebih serius untuk menangani lonjakan tersebut dan mencegah penyebaran penyakit lebih lanjut.

## 4.2 Hasil Standarisasi Data

Tabel 4. 2 Hasil Standarisasi Data

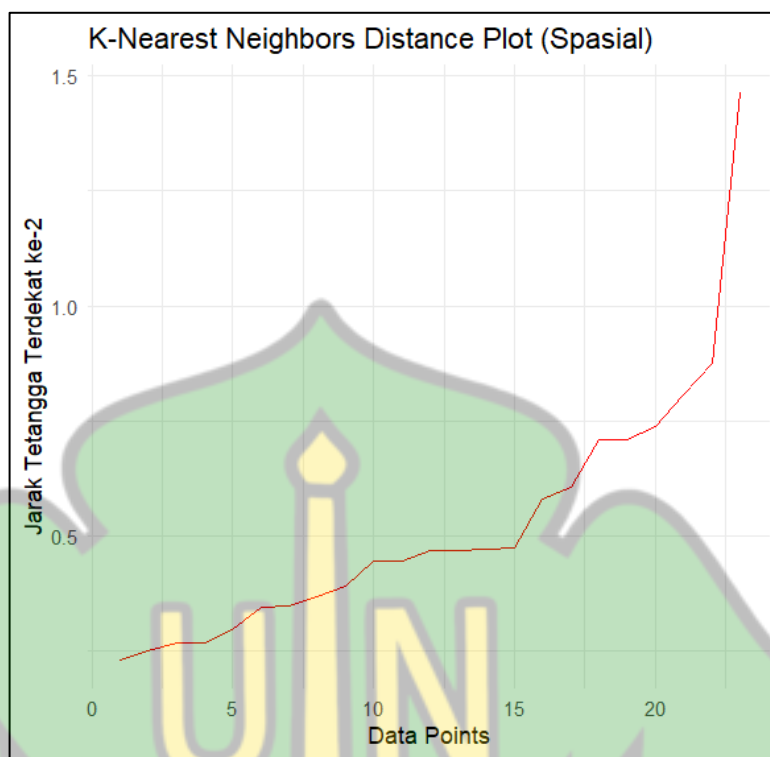
No	Kabupaten/Kota	Hasil Standarisasi
1	Aceh Barat	-0.43113
2	Aceh Barat Daya	-1.27287
3	Aceh Besar	0.052522
4	Aceh Jaya	-0.84494
5	Aceh Selatan	0.516111
6	Aceh Singkil	-0.85608
7	Aceh Tamiang	0.089669
8	Aceh Tengah	-0.63989
9	Aceh Tenggara	-0.63766
10	Aceh Timur	1.123086
11	Aceh Utara	1.826642
12	Banda Aceh	1.730061
13	Bener Meriah	-1.09011
14	Bireuen	1.603763
15	Gayo Lues	-0.7439
16	Langsa	0.020576
17	Lhokseumawe	0.964099

18	Nagan Raya	-0.42444
19	Pidie	1.845215
20	Pidie Jaya	-0.91923
21	Sabang	-1.33676
22	Simeulue	-0.85311
23	Subulussalam	-0.50616

Tabel 4.2 hasil standarisasi menunjukkan variasi jumlah kasus tuberkulosis pada masing-masing kabupaten/kota di Provinsi Aceh. Nilai standarisasi positif mengindikasikan bahwa jumlah kasus TB di suatu wilayah berada di atas nilai rata-rata, sedangkan nilai negatif menunjukkan jumlah kasus berada di bawah rata-rata. Kabupaten Pidie, Aceh Utara, Banda Aceh, dan Bireuen memiliki nilai standarisasi tertinggi, yang menandakan tingkat kasus TB yang relatif tinggi dibandingkan wilayah lainnya. Sebaliknya, Kabupaten Sabang, Aceh Barat Daya, dan Bener Meriah menunjukkan nilai standarisasi terendah, yang mencerminkan jumlah kasus TB yang relatif rendah. Hasil ini menunjukkan adanya ketimpangan distribusi kasus TB antar wilayah di Provinsi Aceh, sehingga diperlukan analisis lanjutan menggunakan metode clustering untuk mengelompokkan wilayah berdasarkan tingkat kesamaan kasus TB.

#### 4.3 Penentuan Parameter Epsilon untuk DBSCAN

Salah satu parameter kunci dalam algoritma DBSCAN adalah *epsilon*, yang menentukan jarak maksimum antara dua titik agar keduanya dapat dianggap sebagai bagian dari satu kluster. Untuk menentukan nilai *epsilon*, digunakan K-Nearest Neighbors Distance Plot, yang menunjukkan jarak ke tetangga terdekat kedua untuk setiap titik data dalam urutan menaik. Pada plot ini, titik "elbow" menandakan nilai *epsilon* yang optimal, yaitu titik di mana terdapat lonjakan signifikan pada jarak antar titik.



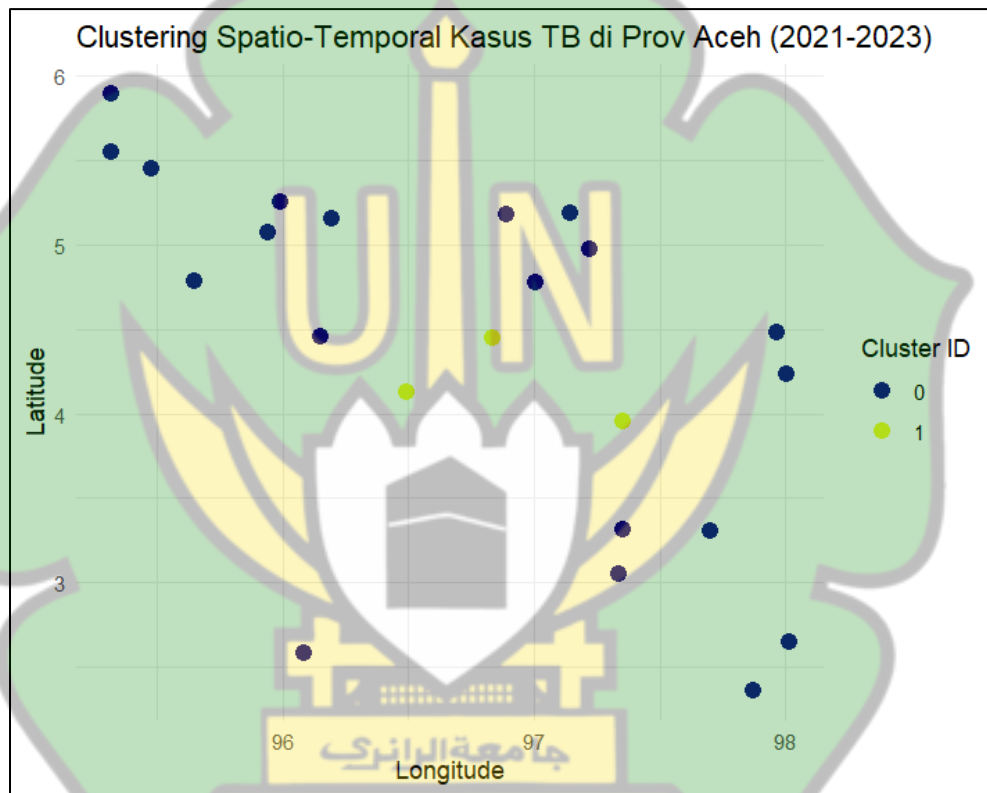
Gambar 4. 2 *K-Nearest Neighbors Distance Plot*

Gambar 4.2 memperlihatkan grafik *K-Nearest Neighbors Distance Plot* yang digunakan sebagai acuan dalam menentukan nilai parameter *epsilon* ( $\epsilon$ ) pada metode klusterisasi berbasis densitas, seperti DBSCAN dan variannya, termasuk DBSCAN. Grafik ini dibentuk berdasarkan data spasial, yaitu koordinat geografis (latitude dan longitude), yang kemudian dihitung jarak ke tetangga ke-2 terdekat ( $k = 2$ ) untuk setiap titik.

Pada sumbu X, ditampilkan urutan titik berdasarkan besar jarak ke tetangga terdekat ke-2, sedangkan sumbu Y menunjukkan nilai dari jarak tersebut. Dari grafik tersebut terlihat adanya titik tekuk (*elbow*) yang cukup mencolok di kisaran nilai  $\epsilon$  antara 0.8 hingga 1.0, yang menjadi indikasi titik optimal untuk pemilihan *epsilon*. Titik elbow ini mencerminkan perubahan signifikan dalam jarak antar titik, yang sering kali dijadikan acuan untuk menentukan batas densitas minimum dalam proses pembentukan kluster.

#### 4.4 Hasil Clustering dengan Metode DBSCAN

Gambar 4.3 menampilkan hasil penerapan algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) pada data spasial kasus Tuberkulosis di Provinsi Aceh selama periode 2019 hingga 2023. Setiap titik pada grafik mewakili lokasi geografis (longitude dan latitude) dari kasus TB yang dianalisis. Berdasarkan nilai parameter *epsilon* dan *MinPts* yang telah ditentukan sebelumnya, DBSCAN berhasil membentuk dua kelompok utama.



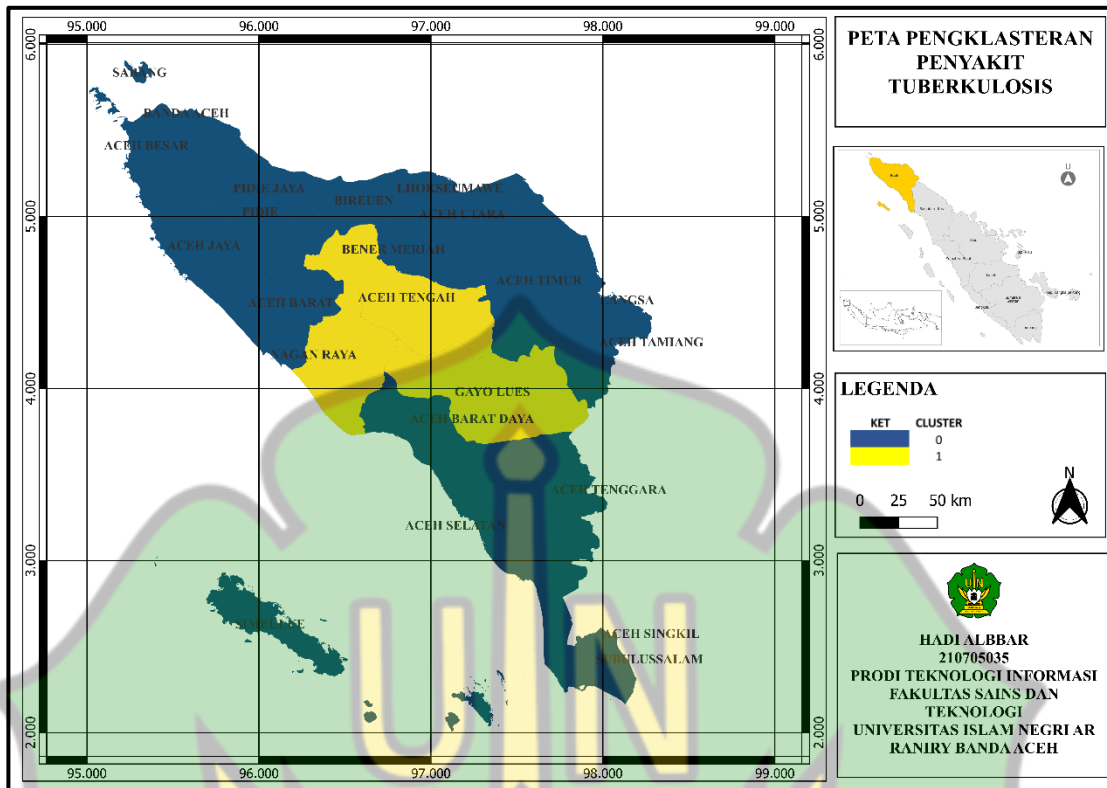
Gambar 4. 3 Hasil Clustering dengan Metode DBSCAN

Hasil pengelompokan data kasus Tuberkulosis (TB) di Provinsi Aceh menggunakan algoritma DBSCAN ditampilkan pada Gambar 4.3. Grafik tersebut menggambarkan hasil proses *clustering* berdasarkan koordinat geografis (longitude dan latitude) yang mewakili posisi masing-masing kabupaten/kota di wilayah penelitian. Setiap titik pada grafik menunjukkan lokasi tertentu, yang kemudian

dikelompokkan berdasarkan tingkat kepadatan data menggunakan prinsip *density-based clustering*. Warna yang berbeda menandakan kelompok (klaster) yang terbentuk, di mana titik-titik dengan warna yang sama termasuk dalam satu kelompok dengan karakteristik kepadatan yang serupa, sedangkan warna yang berbeda menunjukkan klaster yang terpisah. Berdasarkan hasil tersebut, metode DBSCAN berhasil mengidentifikasi dua kelompok utama yang ditandai dengan warna biru dan kuning. Sementara itu, beberapa titik yang berada di luar kedua klaster dapat dikategorikan sebagai *noise* atau wilayah dengan kepadatan data yang terlalu rendah untuk membentuk klaster.

Selanjutnya, hasil pengelompokan tersebut diproyeksikan dalam bentuk peta sebagaimana terlihat pada gambar 4.4 untuk memberikan gambaran spasial dari setiap klaster yang terbentuk. Peta ini menampilkan persebaran wilayah dengan tingkat kepadatan kasus TB yang berbeda di Provinsi Aceh. Daerah berwarna biru merepresentasikan wilayah dengan jumlah kasus yang relatif tinggi dan penyebaran yang lebih padat, sedangkan daerah berwarna kuning menunjukkan wilayah dengan kepadatan kasus yang lebih rendah atau tidak merata.

Visualisasi spasial tersebut memberikan interpretasi yang lebih jelas terhadap hasil *clustering*, karena tidak hanya menunjukkan kelompok data secara numerik, tetapi juga memperlihatkan pola geografis penyebaran penyakit. Hubungan antara gambar 4.3 dan gambar 4.4 menunjukkan bahwa hasil analisis DBSCAN dapat diinterpretasikan secara komprehensif, baik secara statistik maupun spasial. Dengan demikian, pendekatan ini mampu memberikan pemahaman yang lebih mendalam mengenai pola distribusi kasus Tuberkulosis di Provinsi Aceh, yang selanjutnya dapat digunakan sebagai dasar dalam perencanaan kebijakan dan strategi pengendalian penyakit di tingkat wilayah.



Gambar 4. 4 Peta Clustering dengan Metode DBSCAN

Tabel 4. 3 Hasil Clustering dengan Metode DBSCAN

No	Kabupaten/Kota	Cluster
1	Aceh Barat	0
2	Aceh Barat Daya	0
3	Aceh Besar	0
4	Aceh Jaya	0
5	Aceh Selatan	0
6	Aceh Singkil	0
7	Aceh Tamiang	0
8	Aceh Tengah	1
9	Aceh Tenggara	0
10	Aceh Timur	0

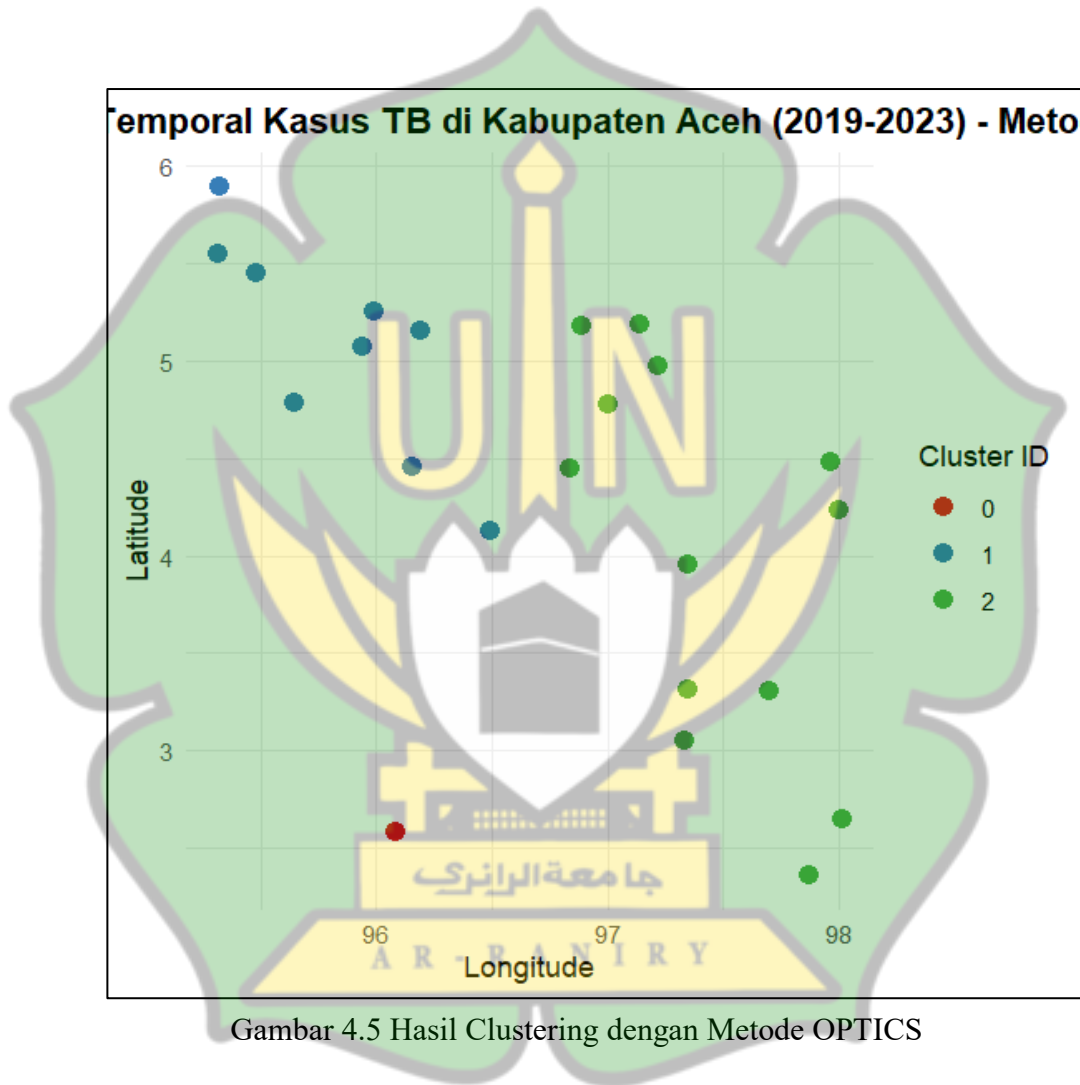
11	Aceh Utara	0
12	Banda Aceh	0
13	Bener Meriah	0
14	Bireuen	0
15	Gayo Lues	1
16	Langsa	0
17	Lhokseumawe	0
18	Nagan Raya	1
19	Pidie	0
20	Pidie Jaya	0
21	Sabang	0
22	Simeulue	0
23	Subulussalam	0

Bedasarkan tabel 4.3 Hasil analisis clustering mengindikasikan bahwa sebagian besar kabupaten di Provinsi Aceh tergolong dalam Cluster 0, yang merepresentasikan kesamaan pola distribusi kasus tuberkulosis dari sisi spasial dan temporal. Sebaliknya, hanya tiga kabupaten yang teridentifikasi berada dalam Cluster 1, yakni Kabupaten Aceh Tengah, Kabupaten Gayo Lues, dan Kabupaten Nagan Raya, yang menunjukkan karakteristik distribusi kasus yang menyimpang secara signifikan dari pola umum yang terbentuk pada wilayah lainnya.

#### 4.5 Hasil Clustering dengan Metode OPTICS

Selain metode DBSCAN, algoritma OPTICS (*Ordering Points To Identify the Clustering Structure*) juga digunakan dalam analisis klastering data spasial-temporal kasus Tuberkulosis (TB). OPTICS merupakan algoritma yang dirancang untuk mengatasi keterbatasan DBSCAN dalam mendeteksi klaster dengan variasi kepadatan data. Berbeda dengan DBSCAN yang mengharuskan penentuan nilai ambang *epsilon* secara tetap, OPTICS mampu secara adaptif mengidentifikasi struktur klaster berdasarkan urutan titik-titik data, sehingga lebih fleksibel dalam mengungkap pola

kluster yang kompleks dan tidak seragam. Keunggulan ini menjadikan OPTICS efektif dalam memetakan distribusi kasus TB yang tersebar secara tidak merata baik dari segi lokasi maupun waktu.



Gambar 4.5 Hasil Clustering dengan Metode OPTICS

Gambar 4.4 menampilkan hasil analisis kluster spasial-temporal kasus Tuberkulosis (TB) di wilayah Provinsi Aceh selama periode 2019 hingga 2023 dengan menggunakan algoritma OPTICS (*Ordering Points To Identify the Clustering Structure*). Setiap titik merepresentasikan satu kabupaten/kota, yang dikelompokkan

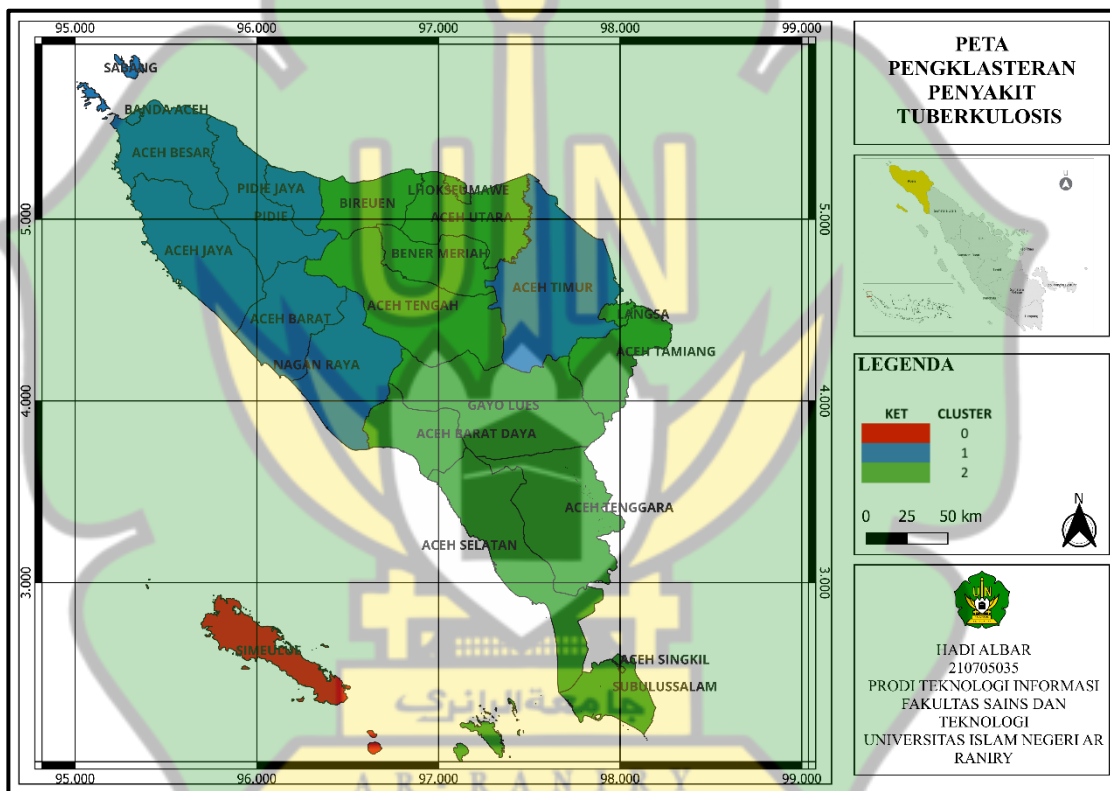
ke dalam tiga klaster berbeda berdasarkan kepadatan dan pola distribusi kasus TB. Warna titik digunakan untuk membedakan masing-masing kluster. Warna biru untuk Cluster 1 menggambarkan wilayah dengan jumlah kasus yang lebih tinggi, menandakan adanya konsentrasi penyebaran TB yang signifikan dan memerlukan penanganan intensif, hijau untuk Cluster 2 encerminkan daerah dengan angka kasus pada tingkat menengah, sehingga penyebaran penyakit berada pada kondisi yang relatif stabil, dan merah untuk Cluster 0 merupakan kelompok penyimpang (*outlier*), yaitu daerah yang jumlah kasusnya berbeda jauh dari pola umum, bisa karena kasus yang sangat rendah ataupun karena faktor geografis tertentu yang khas.

Hasil pemetaan menunjukkan bahwa mayoritas kabupaten/kota di Provinsi Aceh terkelompok ke dalam dua klaster utama, yaitu Cluster 1 dan Cluster 2, yang menandakan penyebaran kasus tuberkulosis cenderung terkonsentrasi pada wilayah tertentu. Sementara itu, terdapat satu kabupaten yang tergolong ke dalam Cluster 0 dan posisinya relatif terpisah dari kelompok lainnya. Kondisi ini mengindikasikan bahwa wilayah tersebut memiliki pola distribusi kasus TB yang berbeda, yang dipengaruhi oleh sejumlah faktor seperti jumlah penduduk yang lebih kecil, letak geografis yang terpencil (misalnya daerah kepulauan), keterbatasan sarana kesehatan, serta perbedaan sistem pencatatan maupun pelaporan kasus.

Sebagai contoh, Kabupaten Simeulue masuk ke dalam Cluster 0 karena memiliki jumlah kasus TB yang jauh lebih rendah dibandingkan daerah lain di Aceh. Lokasinya sebagai wilayah kepulauan membuat penyebaran kasus TB berbeda dari kabupaten/kota di daratan utama, baik dari sisi mobilitas masyarakat, akses layanan kesehatan, maupun kondisi sosial-ekonomi. Faktor-faktor inilah yang menyebabkan Simeulue dikategorikan sebagai *outlier* oleh algoritma OPTICS. Keunggulan OPTICS adalah kemampuannya dalam mendeteksi pola unik seperti ini tanpa harus menggunakan parameter *epsilon* yang tetap, sehingga lebih sesuai diterapkan pada data spasial dengan kepadatan yang bervariasi.”

Visualisasi hasil pengelompokan tersebut kemudian ditampilkan dalam bentuk peta pada ambar 4.6. Peta ini memberikan representasi spasial dari hasil *clustering*

sehingga pola persebaran penyakit TB di Provinsi Aceh dapat diamati secara lebih jelas. Wilayah yang ditandai dengan warna merah menggambarkan daerah dengan kepadatan kasus rendah dan terpisah dari kelompok utama, seperti Kabupaten Simeulue yang termasuk dalam kluster 0. Sementara itu, daerah yang berwarna biru memperlihatkan wilayah dengan tingkat kepadatan kasus menengah, seperti Aceh Selatan, Aceh Barat Daya, dan sekitarnya. Adapun kluster berwarna hijau menggambarkan wilayah dengan kepadatan kasus yang lebih tinggi dan meliputi sebagian besar kabupaten di bagian ujung Barat Provinsi Aceh.



Gambar 4. 6 Peta Clustering dengan Metode OPTICS

Keterkaitan antara Gambar 4.5 dan Gambar 4.6 menunjukkan bahwa hasil analisis *clustering* dengan metode OPTICS tidak hanya dapat dipahami melalui grafik koordinat titik, tetapi juga dapat divisualisasikan secara geografis melalui peta sebaran wilayah. Dengan visualisasi tersebut, pola distribusi penyakit TB dapat diidentifikasi lebih akurat dan informatif. Hal ini membuktikan bahwa metode OPTICS mampu mengelompokkan wilayah berdasarkan perbedaan kepadatan kasus dengan lebih detail serta menggambarkan struktur spasial penyebaran penyakit di Provinsi Aceh secara lebih menyeluruh.

Tabel 4. 4 Hasil *Clustering* dengan Metode OPTICS

No	Kabupaten/Kota	Cluster
1	Aceh Barat	1
2	Aceh Barat Daya	2
3	Aceh Besar	1
4	Aceh Jaya	1
5	Aceh Selatan	2
6	Aceh Singkil	2
7	Aceh Tamiang	2
8	Aceh Tengah	2
9	Aceh Tenggara	2
10	Aceh Timur	1
11	Aceh Utara	2
12	Banda Aceh	1
13	Bener Meriah	2
14	Bireuen	2
15	Gayo Lues	2
16	Langsa	2
17	Lhokseumawe	2
18	Nagan Raya	1
19	Pidie	1

20	Pidie Jaya	1
21	Sabang	1
22	Simeulue	0
23	Subulussalam	1

Berdasarkan table 4.4 hasil analisis kluster terhadap data spasial-temporal kasus Tuberkulosis (TB) Paru di Provinsi Aceh, sebanyak 23 kabupaten/kota terbagi ke dalam tiga kelompok utama. Kelompok pertama (Cluster 1) terdiri dari 9 wilayah, yaitu Aceh Barat, Aceh Besar, Aceh Jaya, Aceh Timur, Banda Aceh, Nagan Raya, Pidie, Pidie Jaya, dan Subulussalam. Wilayah-wilayah ini memiliki pola penyebaran kasus TB yang relatif serupa, yang kemungkinan dipengaruhi oleh faktor seperti kemudahan akses layanan kesehatan, jumlah penduduk, atau kondisi lingkungan yang sebanding, sehingga tingkat kejadian TB di daerah ini tergolong sedang.

Sementara itu, Cluster 2 mencakup 13 kabupaten/kota, yaitu Aceh Barat Daya, Aceh Selatan, Aceh Singkil, Aceh Tamiang, Aceh Tengah, Aceh Tenggara, Aceh Utara, Bener Meriah, Bireuen, Gayo Lues, Langsa, Lhokseumawe, dan Aceh Selatan. Kelompok ini menunjukkan karakteristik yang berbeda, yang mungkin disebabkan oleh kepadatan penduduk yang lebih tinggi, mobilitas masyarakat, atau keterbatasan fasilitas kesehatan, sehingga berpotensi memiliki tingkat penyebaran TB yang lebih kompleks. Sementara itu, Cluster 0 hanya terdiri dari satu wilayah, yakni Kabupaten Simeulue, yang kemungkinan besar memiliki karakteristik penyebaran TB yang unik. Letaknya yang terpisah sebagai wilayah kepulauan menjadikan Simeulue berbeda dari kabupaten/kota lainnya, sehingga masuk dalam kategori outlier. Pengelompokan ini diperoleh melalui algoritma OPTICS, yang mampu mengenali pola kluster dengan kepadatan bervariasi secara otomatis tanpa harus menetapkan parameter *epsilon* secara pasti.

## 4.6 Silhouette Score

Dalam proses analisis kluster, *Silhouette Score* digunakan sebagai metrik untuk mengevaluasi seberapa baik pembentukan kluster oleh suatu algoritma. Skor ini memberikan gambaran mengenai seberapa erat suatu data berada dalam kluster tempatnya tergabung, serta seberapa berbeda jaraknya dibandingkan dengan kluster lainnya. Rentang nilai *Silhouette Score* berada antara -1 hingga 1, di mana skor yang mendekati 1 menunjukkan bahwa objek data berada pada kluster yang sesuai, dengan pemisahan yang jelas dari kluster lain. Sebaliknya, nilai yang mendekati -1 mengindikasikan bahwa objek tersebut lebih menyerupai anggota kluster lain, sehingga kemungkinan ditempatkan pada kluster yang kurang tepat. Nilai ini digunakan untuk menilai kualitas pemisahan kluster secara keseluruhan. Adapun hasil dari perhitungan *Silhouette Score* adalah sebagai berikut:

1. DBSCAN: Tidak dapat menghitung *Silhouette Score* karena hanya ada satu cluster yang valid, yaitu Cluster 0.
2. OPTICS: *Silhouette Score* yang diperoleh dari OPTICS adalah 0.41, yang menunjukkan kualitas clustering yang cukup baik.

#### 4.7 Perbandingan Metode

Tabel 4. 5 Tabel Hasil Perbandingan Metode

Aspek	ST-DBSCAN	OPTICS
<b>Jumlah Cluster</b>	1 Cluster (2 kabupaten) + Banyak <i>Noise</i>	6 Cluster + beberapa <i>Noise</i>
<b>Cluster 1</b>	Aceh Tengah, Gayo Lues, Nagan Raya	Aceh Barat, Aceh Tengah, Aceh Utara, Bireuen, Lhokseumawe, Nagan Raya
<b>Noise</b>	Sebagian besar kabupaten dianggap <i>noise</i>	<i>Noise</i> hanya 1 yaitu kabupaten Simeulue
<b>Fleksibilitas terhadap kepadatan</b>	Kaku (berdasarkan <i>epsilon</i> yang sama)	Fleksibel (mendeteksi cluster dengan kepadatan yang bervariasi)
<b>Deteksi cluster kecil</b>	Tidak sensitif, banyak data dianggap <i>noise</i>	Lebih baik dalam mendeteksi cluster kecil
<b>Kepekaan terhadap perubahan temporal</b>	Lebih sensitif (menggunakan dimensi temporal)	Tidak secara eksplisit menggunakan dimensi temporal
<b><i>Silhouette Score</i></b>	Tidak bisa dihitung (hanya ada satu cluster)	0.41

Penelitian ini mengevaluasi dua algoritma clustering, yakni DBSCAN dan OPTICS, dalam mengidentifikasi pola sebaran kasus tuberkulosis (TB) di wilayah Aceh selama tahun 2019 hingga 2023. Berdasarkan tabel 4.5 hasil analisis, metode OPTICS menunjukkan kinerja yang lebih adaptif terhadap cluster dengan tingkat kepadatan yang bervariasi, serta memperoleh nilai *Silhouette Score* yang lebih tinggi (0,41). Sementara itu, DBSCAN cenderung menghasilkan lebih banyak data yang dianggap *noise*, disebabkan oleh keterbatasannya dalam menangani perbedaan kepadatan antar cluster.

#### 4.8 Analisis Perbedaan Karakteristik Antar Cluster

Setelah proses clustering menggunakan metode OPTICS dan ST-DBSCAN selesai dilakukan, analisis statistik lanjutan dilaksanakan untuk menilai apakah terdapat perbedaan yang signifikan dalam karakteristik kasus Tuberkulosis (TB) antar cluster yang terbentuk. Untuk mengevaluasi hal tersebut, digunakan uji ANOVA guna mengidentifikasi perbedaan perubahan jumlah kasus TB dari tahun 2019 hingga 2023 di antara masing-masing cluster. Apabila hasil uji ANOVA menunjukkan adanya perbedaan yang signifikan secara statistik, maka akan dilanjutkan dengan uji post-hoc Tukey HSD untuk mengetahui secara lebih mendetail perbedaan yang terjadi antar cluster.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(st_dbscan_cluster)	1	631086	631086	1.268	0.273
Residuals	21	10453613	497791		

Gambar 4. 7 Gambar Perbedaan Karakteristik Antar Cluster

Berdasarkan hasil tersebut, diketahui bahwa nilai F sebesar 1.268 dengan nilai signifikansi (p-value) sebesar 0.273. Karena nilai p lebih besar dari tingkat signifikansi 0.05 ( $\alpha = 5\%$ ), maka tidak terdapat perbedaan rata-rata yang signifikan antar cluster yang terbentuk. Dengan kata lain, secara statistik, cluster yang dihasilkan oleh metode DBSCAN belum mampu membedakan wilayah berdasarkan jumlah kasus TB secara signifikan.

Hal ini dapat disebabkan oleh beberapa faktor, seperti karakteristik data yang homogen antar wilayah, atau parameter clustering yang belum optimal dalam memisahkan wilayah dengan beban kasus yang berbeda secara jelas.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan mengenai pemetaan penyebaran kasus Tuberkulosis (TB) di Provinsi Aceh dengan penerapan algoritma *clustering* DBSCAN dan OPTICS, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Metode DBSCAN mampu mengelompokkan wilayah penyebaran TB berdasarkan tingkat kepadatan data yang serupa. Namun, algoritma ini masih menghasilkan sejumlah titik data yang dikategorikan sebagai *noise*, terutama pada wilayah dengan sebaran kasus yang tidak merata. Kelemahan ini muncul karena DBSCAN menggunakan nilai parameter *epsilon* yang tetap, sehingga kurang fleksibel terhadap variasi kepadatan antar wilayah.
2. Metode OPTICS menunjukkan kinerja yang lebih baik dibandingkan DBSCAN dalam mengelompokkan data spasial. OPTICS tidak memerlukan penentuan nilai *epsilon* secara eksplisit, sehingga dapat menyesuaikan diri dengan kepadatan data yang berbeda-beda. Hasil *clustering* menunjukkan adanya tiga kluster utama yang menggambarkan wilayah dengan kepadatan kasus TB rendah, sedang, dan tinggi. Hal ini menandakan bahwa OPTICS lebih adaptif dalam mendeteksi struktur data yang kompleks.
3. Berdasarkan perbandingan hasil *clustering* menggunakan nilai Silhouette Score, metode OPTICS memperoleh hasil yang lebih optimal dengan nilai 0,41, yang menandakan bahwa kualitas pengelompokan cukup baik dan stabil. Dengan demikian, algoritma OPTICS dapat direkomendasikan sebagai metode yang lebih sesuai untuk analisis spasial penyebaran penyakit TB di Provinsi Aceh.
4. Hasil pemetaan yang divisualisasikan dalam bentuk peta digital memberikan gambaran yang lebih jelas mengenai persebaran kasus TB antar kabupaten/kota di Aceh. Peta ini dapat menjadi dasar penting dalam pengambilan keputusan

bagi pihak instansi kesehatan dalam upaya pencegahan dan penanggulangan penyakit TB secara lebih terarah.

## 5.2 Saran

1. Untuk penelitian berikutnya, disarankan agar menggunakan data dengan jangkauan waktu yang lebih panjang serta menambahkan variabel lain seperti faktor sosial, ekonomi, lingkungan, dan perilaku kesehatan masyarakat untuk menghasilkan analisis yang lebih komprehensif.
2. Bagi instansi atau lembaga kesehatan, hasil pemetaan ini diharapkan dapat dimanfaatkan sebagai dasar dalam merumuskan strategi pencegahan dan penanganan kasus TB secara lebih terarah, terutama pada wilayah dengan tingkat kepadatan kasus yang tinggi.
3. Untuk pengembangan metode analisis, penggunaan algoritma lain seperti HDBSCAN atau kombinasi dengan pendekatan berbasis *machine learning* dapat menjadi alternatif untuk meningkatkan akurasi dan mendeteksi pola penyebaran penyakit yang lebih kompleks.
4. Penerapan Sistem Informasi Geografis (SIG) secara terpadu sangat direkomendasikan untuk mendukung visualisasi, pemantauan, dan pengambilan keputusan terkait penyebaran penyakit secara lebih cepat, akurat, dan interaktif.

## DAFTAR PUSTAKA

- Adha, R., Nurhaliza, N., Sholeha, U., & 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. <https://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/12469>
- Amalia, D. (2020). Tingkat Kepatuhan Minum Obat Anti Tuberkulosis Pada Pasien TB Paru Dewasa Rawat Jalan di Puskesmas Dinoyo. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Armiady, D. (2022). Analisis Metode DBSCAN (Density-Based Spatial Clustering of Application with Noise) dalam Mendeteksi Data Outlier. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 2158. <https://doi.org/10.30865/jurikom.v9i6.5080>
- Aryo Wibisono Putra, Zulham Ariyandi, Revi Diwana, Supriyatna, M. U. (2024). Klasterisasi Dan Segmentasi pelanggan Dengan Menggunakan Algoritma Ordering Points To Identify the Clustering Structure (Optics) Dan Dbscan. *SENASTIKA Universitas Malikussaleh KLAUSTERISASI*.
- Asyuti, S., & Setyawan, A. A. (2023). Data Mining Dalam Penggunaan Presensi Karyawan Dengan Cluster Means. *Jurnal Ilmiah Sains Teknologi Dan Informasi*, 1(1), 01–10. <https://jurnal.alimspublishing.co.id/index.php/JITI/article/download/6/6>
- Aulia Permatasari, N., Herry Chrisnanto, Y., Kania Ningsih, A., Terusan Jendral Sudirman, J., Selatan, C., & Jawa Barat, C. (2023). Segmentasi Kasus Data Kematian Covid 19 Di Jawa Barat Menggunakan Algoritma DBSCAN Universitas Jendral Achmad Yani. *IJESPG Journal*, 1(4), 119–128. <http://ijespgjournal.org>
- Di, T., & Tunggal, D. (2023). *Jurnal Pengabdian Komunitas*. 02(02), 108–113.
- Dinas Kesehatan Kota Surabaya. (2017). Profil Dinas Kesehatan Kota Surabaya. *Dinas Kesehatan*, 163.
- Fadilah, Z. R., & Wijayanto, A. W. (2023). Perbandingan Metode Klasterisasi Data Bertipe Campuran: One-Hot-Encoding, Gower Distance, dan K-Prototype Berdasarkan Akurasi (Studi Kasus: Chronic Kidney Disease Dataset). *Journal of Applied Informatics and Computing*, 7(1), 57–67. <https://doi.org/10.30871/jaic.v7i1.5857>

- Fitri, E. M., Suryono, R. R., & Wantoro, A. (2023). Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz. *Jurnal Komputasi*, 11(2), 157–168. <https://doi.org/10.23960/komputasi.v11i2.12582>
- Hastuti, S. H., Septiani, A., Hendrayani, H., & Nurmayanti, W. P. (2024). Penerapan Metode OPTICS dan ST-DBSCAN untuk Klasterisasi Data Kesehatan. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 252–261. <https://doi.org/10.29408/edumatic.v8i1.25765>
- Indini, D. P., Siburian, S. R., Utomo, D. P., Studi, P., Informatika, T., Darma, U. B., Non-akademik, P., & Mining, D. (2022). *Implementasi Algoritma Dbscan Untuk Clustering Seleksi*. 1328–1335.
- Juardi, M. I. (2025). *Perbandingan algoritma k-means dan dbscan pada clustering makanan berdasarkan kandungan nutrisi berbasis web skripsi*.
- Kanda, R. L., & Tanggo, W. D. (2022). Program studi sarjana keperawatan dan ners sekolah tinggi kesehatan stella maris makassar 2022. *Jurnal Stella Maris Makassar 2022*, 10–80.
- Kodariah, L., Murtafi'ah, N., & Baehaki, F. (2023). Pemeriksaan dan peningkatan pengetahuan masyarakat tentang penularan penyakit tuberculosis dalam rangka menurunkan angka penularan. *Jurnal Inovasi Hasil Pengabdian Masyarakat (JIPEMAS)*, 6(1), 59–70. <https://doi.org/10.33474/jipemas.v6i1.17971>
- Mustikasari, & Salman, N. (2023). Analisis kluster berbasis kepadatan dengan dbscan dan optics. *Insypro Information System and Processing*, 8(1), 1–8. <http://journal.uinalauddin.ac.id/index.php/insypro>
- Nisrina, S., Nurmayanti, W. P., Basirun, Kertanah, & Muhammad Gazali. (2022). Penerapan Metode Clustering SOM dan DBSCAN dalam Mengelompokkan Unmet Need Keluarga Berencana di Nusa Tenggara Barat. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(2), 237–244. <https://doi.org/10.36456/jstat.vol15.no2.a5549>
- Novi, S. (2020). *Clustering Status Desa Menggunakan Metode Dbscan (Density-Based Spatialclustering of Application With Noise)*. xi–xii.
- Nurhalizah, C. D., Damaliana, A. T., & Prasetya, D. A. (2025). *OPTICS-Based Clustering of East Java Regencies and Cities by Divorce Factors*. 4(3), 113–123.
- Okfalisa, O., Angraini, A., Novi, S., Rusnedy, H., Handayani, L., & Mustakim, M.

- (2021). Identification of the distribution village maturation: Village classification using Density-based spatial clustering of applications with noise. *Jurnal Teknologi Dan Sistem Komputer*, 9(3), 133–141. <https://doi.org/10.14710/jtsiskom.2021.13998>
- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48. [https://doi.org/10.51557/pt\\_jiit.v6i2.659](https://doi.org/10.51557/pt_jiit.v6i2.659)
- Pakuani, K. W., & Kurniawan, R. (2021). Kajian Penentuan Nilai Epsilon Optimal Pada Algoritma DMDBSCAN Dan Pemetaan Daerah Rawan Gempa Bumi Di Indonesia Tahun 2014-2020. *Seminar Nasional Official Statistics*, 2021(1), 991–1000. <https://doi.org/10.34123/semnasoffstat.v2021i1.847>
- Pokhrel, S. (2024). No TitleEΛENH. *Aγαη*, 15(1), 37–48.
- Puspitasari, D. A., Cahyana, Y., & Lestari, S. A. P. (2023). Penerapan Algoritma Density Based Spastial Clustering Algorithm With Noise Untuk Pengelompokkan Penyakit Pasien. *Scientific Student Journal for Information, Technology and Science*, 4(1), 102–106.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Saputra, A., & Yusuf, R. (2024). Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346–1361. <https://doi.org/10.57152/malcom.v4i4.1516>
- Sari, D. P., Rizqi, T. A., Fitrianto, A., Erfiani, M. S., & Jumansyah, R. D. (2024). *Perbandingan Metode K-Means dan OPTICS dalam Penggerombolan Kemiskinan Multidimensi di Indonesia*. 198–207.
- Setiyawan, D. T., & Barkah, A. S. (2025). *Comparative Analysis of DBSCAN , OPTICS , and Agglomerative Clustering Methods for Identifying Disease Distribution Patterns in Banjarnegara Community Health Centers*. 6(3), 1229–1239.
- Setyani, T., Nopijani Heidy, H., Sari, K., Indriani, Y., & Sulistiani, H. (2025a). Perbandingan Kinerja Algoritma K-Means dan DBSCAN dalam Segmentasi

- Nasabah Berdasarkan Data Pemasaran Bank. *Prosiding Seminar Nasional Teknologi Komputer Dan Sains*, 3(1), 376–382.  
<https://doi.org/10.47065/mis.v2i3>.
- Setyani, T., Nopijani Heidy, H., Sari, K., Indriani, Y., & Sulistiani, H. (2025b). Perbandingan Kinerja Algoritma K-Means dan DBSCAN dalam Segmentasi Nasabah Berdasarkan Data Pemasaran Bank. *Prosiding Seminar Nasional Teknologi Komputer Dan Sains*, 3(1), 376–382.  
<https://prosiding.seminars.id/sainteks>
- Simbolon, I. N., & Friskila, P. D. (2024). Analisis Dan Evaluasi Algoritma Dbscan (Density-Based Spatial Clustering of Applications With Noise) Pada Tuberkulosis. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3S1).  
<https://doi.org/10.23960/jitet.v12i3s1.5206>
- Sujatmiko, B. (2023). *Analisis Metode Clustering dengan Algoritma Density-Based Spatial Clustering of Application With Noise (DBSCAN) untuk Pengelompokan Cyberbullying di Instagram*.
- Sulistyo, A. ; W. (2020). *Implementasi Metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dalam Mengelompokkan Penyebaran Tuberkulosis (Studi Kasus: Kluster Data Penyebaran Tuberkulosis High Burden Country (HBC) pada Tahun 2018)*.
- Syawali, Y., Haikal Hafiz Rangkuti, M., Aprianda Mayadi, K., Silvi Pratiwi, K., Fahrur Razi, M., Ilmu Komputer, J., & Matematika dan Ilmu Pengetahuan Alam, F. (2024). PENERAPAN ALGORITMA DBSCAN UNTUK ANALISIS DEMOGRAFIS dan PENGELUARAN PELANGGAN MALL. *Jurnal Matematika Sains Sahombu*.  
<https://jumas.ourhope.biz.id/ojs/index.php/JM/article/download/27/14>
- Talud, K. L., Wattimanela, H. J., Laamena, N. S., & Matdoan, M. Y. (2025). Implementasi Algoritma Density Based Spatial Clustering of Application With Noise (DBSCAN) Studi Indeks Pembangunan Manusia. *Codeverse: The Journal of Emerging DigitalRealitie*, 01(1), 9–20.
- TURGUT, E., TAŞYÜREK, M., & AZGINOĞLU, N. (2022). Detection of Buildings in the Urban Transformation Process with Spatial Data Mining Methods. *DÜMF Mühendislik Dergisi*, 2, 161–167. <https://doi.org/10.24012/dumf.1072110>
- Ummah, M. S. (2019). No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title. *Sustainability (Switzerland)*, 11(1),

1–14. [http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbe.co.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484\\_SISTEM\\_PEMBETUNGAN\\_TERPUSAT\\_STRATEGI\\_MELESTARI](http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbe.co.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI)

Wijaya, T. A., Utami, E., & Al Fatta, H. (2024). Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Data Gangguan PT. PLN UID Kalselteng. *Innovative: Journal Of Social Science Research*, 4(1), 8846–8854.

Zhafar, M. N., Usman, K., Akhyar, F., Lashiyanti, A. R., Rasyid Munthe, I., Nasution, F. A., Korespondensi, E. P., Huda Aminuddin, F., Djauhari, T., Arjansyah, A., Norshahlan, M., Jaya, H., Kustini, R., Nurahman, N., Purwanto, A., Mulyanto, S., Mutaqin, H., Guntara, M., Lutfi, N., ... Nurina Sari, B. (2023). Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Analisa Persebaran Varian Covid-19 ( Studi Kasus Kelurahan Antapani Kidul ). *E-Proceeding of Engineering*, 2(1), 1042. <http://jurnal.mdp.ac.id>

Adha, R., Nurhaliza, N., Sholeha, U., & 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. <https://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/12469>

Amalia, D. (2020). Tingkat Kepatuhan Minum Obat Anti Tuberkulosis Pada Pasien TB Paru Dewasa Rawat Jalan di Puskesmas Dinoyo. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.

Armiady, D. (2022). Analisis Metode DBSCAN (Density-Based Spatial Clustering of Application with Noise) dalam Mendeteksi Data Outlier. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 2158. <https://doi.org/10.30865/jurikom.v9i6.5080>

Aryo Wibisono Putra, Zulham Ariyandi, Revi Diwana, Supriyatna, M. U. (2024). Klasterisasi Dan Segmentasi Pelanggan Dengan Menggunakan Algoritma Ordering Points To Identify the Clustering Structure (Optics) Dan Dbscan. *SENASTIKA Universitas Malikussaleh KLASTERISASI*.

Asyuti, S., & Setyawan, A. A. (2023). Data Mining Dalam Penggunaan Presensi Karyawan Denga Cluster Means. *Jurnal Ilmiah Sains Teknologi Dan Informasi*, 1(1), 01–10. <https://jurnal.alimspublishing.co.id/index.php/JITI/article/download/6/6>

Aulia Permatasari, N., Herry Chrisnanto, Y., Kania Ningsih, A., Terusan Jendral

- Sudirman, J., Selatan, C., & Jawa Barat, C. (2023). Segmentasi Kasus Data Kematian Covid 19 Di Jawa Barat Menggunakan Algoritma DBSCAN Universitas Jendral Achmad Yani. *IJESPG Journal*, 1(4), 119–128. <http://ijespgjournal.org>
- Di, T., & Tunggul, D. (2023). *Jurnal Pengabdian Komunitas*. 02(02), 108–113.
- Dinas Kesehatan Kota Surabaya. (2017). Profil Dinas Kesehatan Kota Surabaya. *Dinas Kesehatan*, 163.
- Fadilah, Z. R., & Wijayanto, A. W. (2023). Perbandingan Metode Klasterisasi Data Bertipe Campuran: One-Hot-Encoding, Gower Distance, dan K-Prototype Berdasarkan Akurasi (Studi Kasus: Chronic Kidney Disease Dataset). *Journal of Applied Informatics and Computing*, 7(1), 57–67. <https://doi.org/10.30871/jaic.v7i1.5857>
- Fitri, E. M., Suryono, R. R., & Wantoro, A. (2023). Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz. *Jurnal Komputasi*, 11(2), 157–168. <https://doi.org/10.23960/komputasi.v11i2.12582>
- Hastuti, S. H., Septiani, A., Hendrayani, H., & Nurmayanti, W. P. (2024). Penerapan Metode OPTICS dan ST-DBSCAN untuk Klasterisasi Data Kesehatan. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 252–261. <https://doi.org/10.29408/edumatic.v8i1.25765>
- Indini, D. P., Siburian, S. R., Utomo, D. P., Studi, P., Informatika, T., Darma, U. B., Non-akademik, P., & Mining, D. (2022). *Implementasi Algoritma Dbscan Untuk Clustering Seleksi*. 1328–1335.
- Juardi, M. I. (2025). *Perbandingan algoritma k-means dan dbscan pada clustering makanan berdasarkan kandungan nutrisi berbasis web skripsi*.
- Kanda, R. L., & Tanggo, W. D. (2022). Program studi sarjana keperawatan dan ners sekolah tinggi kesehatan stella maris makassar 2022. *Jurnal Stella Maris Makassar 2022*, 10–80.
- Kodariah, L., Murtafi'ah, N., & Baehaki, F. (2023). Pemeriksaan dan peningkatan pengetahuan masyarakat tentang penularan penyakit tuberculosis dalam rangka menurunkan angka penularan. *Jurnal Inovasi Hasil Pengabdian Masyarakat (JIPEMAS)*, 6(1), 59–70. <https://doi.org/10.33474/jipemas.v6i1.17971>
- Mustikasari, & Salman, N. (2023). Analisis kluster berbasis kepadatan dengan dbscan dan optics. *Insypro Information System and Processing*, 8(1), 1–8.

<http://journal.uinalauddin.ac.id/index.php/insypro>

- Nisrina, S., Nurmayanti, W. P., Basirun, Kertanah, & Muhammad Gazali. (2022). Penerapan Metode Clustering SOM dan DBSCAN dalam Mengelompokkan Unmet Need Keluarga Berencana di Nusa Tenggara Barat. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(2), 237–244. <https://doi.org/10.36456/jstat.vol15.no2.a5549>
- Novi, S. (2020). *Clustering Status Desa Menggunakan Metode Dbscan (Density-Based Spatialclustering of Application With Noise)*. xi–xii.
- Nurhalizah, C. D., Damaliana, A. T., & Prasetya, D. A. (2025). *OPTICS-Based Clustering of East Java Regencies and Cities by Divorce Factors*. 4(3), 113–123.
- Okfalisa, O., Angraini, A., Novi, S., Rusnedy, H., Handayani, L., & Mustakim, M. (2021). Identification of the distribution village maturation: Village classification using Density-based spatial clustering of applications with noise. *Jurnal Teknologi Dan Sistem Komputer*, 9(3), 133–141. <https://doi.org/10.14710/jtsiskom.2021.13998>
- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48. [https://doi.org/10.51557/pt\\_jiit.v6i2.659](https://doi.org/10.51557/pt_jiit.v6i2.659)
- Pakuani, K. W., & Kurniawan, R. (2021). Kajian Penentuan Nilai Epsilon Optimal Pada Algoritma DMDBSCAN Dan Pemetaan Daerah Rawan Gempa Bumi Di Indonesia Tahun 2014–2020. *Seminar Nasional Official Statistics, 2021*(1), 991–1000. <https://doi.org/10.34123/semnasoffstat.v2021i1.847>
- Pokhrel, S. (2024). No TitleΕΛΕΝΗ. *Αγαν*, 15(1), 37–48.
- Puspitasari, D. A., Cahyana, Y., & Lestari, S. A. P. (2023). Penerapan Algoritma Density Based Spastial Clustering Algorithm With Noise Untuk Pengelompokkan Penyakit Pasien. *Scientific Student Journal for Information, Technology and Science*, 4(1), 102–106.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Saputra, A., & Yusuf, R. (2024). Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta

- Menggunakan Metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346–1361.  
<https://doi.org/10.57152/malcom.v4i4.1516>
- Sari, D. P., Rizqi, T. A., Fitrianto, A., Erfiani, M. S., & Jumansyah, R. D. (2024). *Perbandingan Metode K-Means dan OPTICS dalam Penggerombolan Kemiskinan Multidimensi di Indonesia*. 198–207.
- Setiyawan, D. T., & Barkah, A. S. (2025). *Comparative Analysis of DBSCAN, OPTICS, and Agglomerative Clustering Methods for Identifying Disease Distribution Patterns in Banjarnegara Community Health Centers*. 6(3), 1229–1239.
- Setyani, T., Nopijani Heidy, H., Sari, K., Indriani, Y., & Sulistiani, H. (2025a). Perbandingan Kinerja Algoritma K-Means dan DBSCAN dalam Segmentasi Nasabah Berdasarkan Data Pemasaran Bank. *Prosiding Seminar Nasional Teknologi Komputer Dan Sains*, 3(1), 376–382.  
<https://doi.org/10.47065/mis.v2i3>.
- Setyani, T., Nopijani Heidy, H., Sari, K., Indriani, Y., & Sulistiani, H. (2025b). Perbandingan Kinerja Algoritma K-Means dan DBSCAN dalam Segmentasi Nasabah Berdasarkan Data Pemasaran Bank. *Prosiding Seminar Nasional Teknologi Komputer Dan Sains*, 3(1), 376–382.  
<https://prosiding.seminars.id/sainteks>
- Simbolon, I. N., & Friskila, P. D. (2024). Analisis Dan Evaluasi Algoritma Dbscan (Density-Based Spatial Clustering of Applications With Noise) Pada Tuberkulosis. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3S1).  
<https://doi.org/10.23960/jitet.v12i3s1.5206>
- Sujatmiko, B. (2023). *Analisis Metode Clustering dengan Algoritma Density-Based Spatial Clustering of Application With Noise (DBSCAN) untuk Pengelompokan Cyberbullying di Instagram*.
- Sulistyo, A. ; W. (2020). *Implementasi Metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN) dalam Mengelompokkan Penyebaran Tuberkulosis (Studi Kasus: Klaster Data Penyebaran Tuberkulosis High Burden Country (HBC) pada Tahun 2018)*.
- Syawali, Y., Haikal Hafiz Rangkuti, M., Aprianda Mayadi, K., Silvi Pratiwi, K., Fahrur Razi, M., Ilmu Komputer, J., & Matematika dan Ilmu Pengetahuan Alam, F. (2024). PENERAPAN ALGORITMA DBSCAN UNTUK ANALISIS

DEMOGRAFIS dan PENGELUARAN PELANGGAN MALL. *Jurnal Matematika Sains Sahombu*.

<https://jumas.ourhope.biz.id/ojs/index.php/JM/article/download/27/14>

- Talud, K. L., Wattimanela, H. J., Laamena, N. S., & Matdoan, M. Y. (2025). Implementasi Algoritma Density Based Spatial Clustering of Application With Noise (DBSCAN) Studi Indeks Pembangunan Manusia. *Codeverse: The Journal of Emerging Digital Realities*, 01(1), 9–20.
- TURGUT, E., TAŞYÜREK, M., & AZGINOĞLU, N. (2022). Detection of Buildings in the Urban Transformation Process with Spatial Data Mining Methods. *DÜMF Mühendislik Dergisi*, 2, 161–167. <https://doi.org/10.24012/dumf.1072110>
- Ummah, M. S. (2019). No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析 Title. *Sustainability (Switzerland)*, 11(1), 1–14. [http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbe.co.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484\\_SISTEM\\_PEMBETUNGAN\\_TERPUSAT\\_STRATEGI\\_MELESTARI](http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbe.co.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI)
- Wijaya, T. A., Utami, E., & Al Fatta, H. (2024). Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Data Gangguan PT. PLN UID Kalselteng. *Innovative: Journal Of Social Science Research*, 4(1), 8846–8854.
- Zhafar, M. N., Usman, K., Akhyar, F., Lashiyanti, A. R., Rasyid Munthe, I., Nasution, F. A., Korespondensi, E. P., Huda Aminuddin, F., Djauhari, T., Arjansyah, A., Norshahlan, M., Jaya, H., Kustini, R., Nurahman, N., Purwanto, A., Mulyanto, S., Mutaqin, H., Guntara, M., Lutfi, N., ... Nurina Sari, B. (2023). Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Analisa Persebaran Varian Covid-19 ( Studi Kasus Kelurahan Antapani Kidul ). *E-Proceeding of Engineering*, 2(1), 1042. <http://jurnal.mdp.ac.id>

## LAMPIRAN

```
> # Load the necessary libraries
> library(dbscan)
> library(factoextra)
```

Welcome! want to learn more? See two factoextra-related books at <https://goo.gl/ve3WBa>

Warning message:

```
package 'factoextra' was built under R version 4.3.3
> library(dplyr)
> library(ggplot2)
> library(multcompView)
```

Warning message:

```
package 'multcompView' was built under R version 4.3.3
> library(FNN)
```

Warning message:

```
package 'FNN' was built under R version 4.3.3
> library(cluster)
```

Warning message:

```
package 'cluster' was built under R version 4.3.3
> # Membuat data kasus TB di Aceh untuk tahun 2022 dan 2023
> tb_data <- data.frame(
+   Kabupaten = c('Aceh Barat', 'Aceh Barat Daya', 'Aceh Besar', 'Aceh Jaya', 'Aceh Selatan',
+                 'Aceh Singkil', 'Aceh Tamiang', 'Aceh Tengah', 'Aceh Tenggara', 'Aceh Timur',
+                 'Aceh Utara', 'Banda Aceh', 'Bener Meriah', 'Bireuen', 'Gayo Lues',
+                 'Langsa', 'Lhokseumawe', 'Nagan Raya', 'Pidie', 'Pidie Jaya',
+                 'Sabang', 'Simeulue', 'Subulussalam'),
+   TB_2019 = c(311, 202, 408, 212, 374, 208, 426, 180, 129, 567, 901, 997, 67, 848, 199, 410,
+               598, 284, 799, 199, 30, 124, 174),
+   TB_2020 = c(259, 232, 347, 155, 281, 95, 259, 189, 58, 578, 762, 379, 100, 740, 105, 249,
+               293, 123, 587, 65, 18, 149, 75),
+   TB_2021 = c(184, 247, 355, 120, 415, 67, 447, 147, 160, 648, 671, 604, 62, 729, 149,
+               304, 587, 180, 718, 85, 9, 80, 202),
+   TB_2022 = c(290, 262, 437, 143, 747, 196, 451, 266, 345, 813, 1016, 1108, 104, 876,
+               226, 485, 853, 373, 1144, 161, 28, 201, 371),
+   TB_2023 = c(291, 315, 439, 148, 793, 197, 453, 272, 365, 821, 1024, 1156, 115, 881,
```

```

+           235, 495, 882, 384, 1151, 168, 31, 213, 415),
+ Latitude = c(4.4542745, 3.0512643, 5.4529168, 4.7873684, 3.31150
56,
+           2.3589459, 4.2328871, 4.4482641, 3.3088666, 5.25544
3,
+           4.9786331, 5.55, 4.7748348, 5.18254, 3.955165,
+           4.48, 5.1880556, 4.1248406, 5.0742659, 5.1548063,
+           5.8946929, 2.583333, 2.6449927),
+ Longitude = c(96.1526985, 97.3368031, 95.4777811, 95.6457951, 97
.3516558,
+           97.87216, 98.0028892, 96.8350999, 97.6982272, 95.9
885456,
+           97.2221421, 95.3166667, 97.0068393, 96.89005, 97.3
516558,
+           97.9633333, 97.1402778, 96.4929797, 95.940971, 96.
195132,
+           95.3192982, 96.083333, 98.0165205)
+ )
> # Menghitung perubahan jumlah kasus TB (Delta TB)
> tb_data$Delta_TB <- tb_data$TB_2023 - tb_data$TB_2022 - tb_data$TB
_2021 -
+ tb_data$TB_2020 - tb_data$TB_2019
> print(tb_data$Delta_TB)
 [1] -753 -628 -1108 -482 -1024 -369 -1130 -510 -327 -1785 -23
26 -1932 -218 -2312 -444 -953 -1449
 [18] -576 -2097 -342 -54 -341 -407
> # Menggunakan pivot_longer dari tidyr untuk mengubah format data
> library(tidyr)
Warning message:
package 'tidyr' was built under R version 4.3.3
> # Mengubah data dari wide ke long menggunakan pivot_longer
> tb_long <- tb_data %>%
+   pivot_longer(cols = c(TB_2019, TB_2020, TB_2021, TB_2022, TB_202
3),
+               names_to = "Year",
+               values_to = "Cases")
> # Membuat boxplot untuk kasus TB 2019 dan 2023
> library(ggplot2)
> ggplot(tb_long, aes(x = Year, y = Cases)) +
+   geom_boxplot(fill = "lightblue") +
+   labs(title = "Boxplot Kasus TB di Aceh (2019 - 2023)", x = "Tahu
n", y = "Jumlah Kasus TB") +
+   theme_minimal()
> #EPSILON
> library(FNN) # Paket untuk KNN
> library(ggplot2)
> # Mengambil data spasial (latitude dan longitude)
> spatial_data_kabupaten <- tb_data %>% select(Latitude, Longitude)
> # Menggunakan KNN untuk mencari tetangga terdekat
> knn_result <- get.knn(spatial_data_kabupaten, k = 2) # Tetangga t
erdekat ke-2

```

```

> # Mengambil jarak tetangga terdekat ke-2 (kolom kedua dari hasil)
> distances <- sort(knn_result$nn.dist[, 2])
> # Membuat plot jarak tetangga terdekat untuk mencari titik "elbow"
> df_distances <- data.frame(Index = 1:length(distances), Distance =
distances)
> ggplot(df_distances, aes(x = Index, y = Distance)) +
+   geom_line(color = "red") +
+   labs(title = "K-Nearest Neighbors Distance Plot (Spasial)",
+         x = "Data Points", y = "Jarak Tetangga Terdekat ke-2") +
+   theme_minimal()
> #STDBSCAN
> library(dbSCAN)
> library(ggplot2)
> library(dplyr)
> library(cluster)
> # Menggabungkan data spasial (latitude, longitude) dan data tempor
al (perubahan kasus TB)
> spatial_data_kabupaten <- tb_data %>% select(Latitude, Longitude)
> temporal_data_kabupaten <- tb_data %>% mutate(delta_TB_2019_2020_2
021_2022_2023 = Delta_TB / 100) # Normalisasi data temporal
> temporal_data_kabupaten
  Kabupaten TB_2019 TB_2020 TB_2021 TB_2022 TB_2023 Latitude
Longitude Delta_TB
1   Aceh Barat    311    259    184    290    291 4.454275
96.15270   -753
2 Aceh Barat Daya  202    232    247    262    315 3.051264
97.33680   -628
3   Aceh Besar    408    347    355    437    439 5.452917
95.47778  -1108
4   Aceh Jaya     212    155    120    143    148 4.787368
95.64580   -482
5   Aceh Selatan  374    281    415    747    793 3.311506
97.35166  -1024
6   Aceh Singkil  208     95     67    196    197 2.358946
97.87216   -369
7   Aceh Tamiang  426    259    447    451    453 4.232887
98.00289  -1130
8   Aceh Tengah   180    189    147    266    272 4.448264
96.83510   -510
9   Aceh Tenggara 129     58    160    345    365 3.308867
97.69823   -327
10  Aceh Timur    567    578    648    813    821 5.255443
95.98855  -1785
11  Aceh Utara     901    762    671    1016   1024 4.978633
97.22214  -2326
12  Banda Aceh    997    379    604    1108   1156 5.550000
95.31667  -1932
13  Bener Meriah   67    100     62    104    115 4.774835
97.00684   -218
14  Bireuen       848    740    729    876    881 5.182540
96.89005  -2312

```

15	Gayo Lues	199	105	149	226	235	3.955165
	97.35166	-444					
16	Langsa	410	249	304	485	495	4.480000
	97.96333	-953					
17	Lhokseumawe	598	293	587	853	882	5.188056
	97.14028	-1449					
18	Nagan Raya	284	123	180	373	384	4.124841
	96.49298	-576					
19	Pidie	799	587	718	1144	1151	5.074266
	95.94097	-2097					
20	Pidie Jaya	199	65	85	161	168	5.154806
	96.19513	-342					
21	Sabang	30	18	9	28	31	5.894693
	95.31930	-54					
22	Simeulue	124	149	80	201	213	2.583333
	96.08333	-341					
23	Subulussalam	174	75	202	371	415	2.644993
	98.01652	-407					

delta\_TB\_2019\_2020\_2021\_2022\_2023

1	-7.53
2	-6.28
3	-11.08
4	-4.82
5	-10.24
6	-3.69
7	-11.30
8	-5.10
9	-3.27
10	-17.85
11	-23.26
12	-19.32
13	-2.18
14	-23.12
15	-4.44
16	-9.53
17	-14.49
18	-5.76
19	-20.97
20	-3.42
21	-0.54
22	-3.41
23	-4.07

```

> # Gabungkan data spasial dan temporal menjadi satu matriks
> st_data_kabupaten <- cbind(spatial_data_kabupaten, temporal_data_kabupaten$delta_TB_2019_2020_2021_2022_2023)
> # Terapkan ST-DBSCAN dengan parameter epsilon yang sesuai
> st_dbscan_model <- dbSCAN(st_data_kabupaten, eps = 1, minPts = 3)
> # Simpan hasil clustering ke dalam dataframe tb_data
> tb_data$st_dbscan_cluster <- st_dbscan_model$cluster
> tb_data$st_dbscan_cluster <- st_dbscan_model$cluster + 1
> # visualisasi hasil clustering menggunakan ggplot2

```

```

> ggplot(tb_data, aes(x = Longitude, y = Latitude, color = factor(st
_dbscan_cluster))) +
+   geom_point(size = 3) +
+   scale_color_viridis_d(option = "plasma") +
+   labs(title = "Clustering Spatio-Temporal Kasus TB di Prov Aceh (
2021-2023)",
+         x = "Longitude", y = "Latitude", color = "Cluster ID") +
+   theme_minimal()
> # Menampilkan hasil cluster dalam bentuk tabel
> hasil_cluster <- tb_data %>%
+   select(Kabupaten, Latitude, Longitude, TB_2023, TB_2022, TB_2021
, TB_2020, TB_2019, Delta_TB, st_dbscan_cluster)
> print(hasil_cluster)

```

	Kabupaten	Latitude	Longitude	TB_2023	TB_2022	TB_2021	TB_2020
0	TB_2019	Delta_TB	st_dbscan_cluster				
1	Aceh Barat	4.454275	96.15270	291	290	184	25
9	311	-753	1				
2	Aceh Barat Daya	3.051264	97.33680	315	262	247	23
2	202	-628	1				
3	Aceh Besar	5.452917	95.47778	439	437	355	34
7	408	-1108	1				
4	Aceh Jaya	4.787368	95.64580	148	143	120	15
5	212	-482	1				
5	Aceh Selatan	3.311506	97.35166	793	747	415	28
1	374	-1024	1				
6	Aceh Singkil	2.358946	97.87216	197	196	67	9
5	208	-369	1				
7	Aceh Tamiang	4.232887	98.00289	453	451	447	25
9	426	-1130	1				
8	Aceh Tengah	4.448264	96.83510	272	266	147	18
9	180	-510	2				
9	Aceh Tenggara	3.308867	97.69823	365	345	160	5
8	129	-327	1				
10	Aceh Timur	5.255443	95.98855	821	813	648	57
8	567	-1785	1				
11	Aceh Utara	4.978633	97.22214	1024	1016	671	76
2	901	-2326	1				
12	Banda Aceh	5.550000	95.31667	1156	1108	604	37
9	997	-1932	1				
13	Bener Meriah	4.774835	97.00684	115	104	62	10
0	67	-218	1				
14	Bireuen	5.182540	96.89005	881	876	729	74
0	848	-2312	1				
15	Gayo Lues	3.955165	97.35166	235	226	149	10
5	199	-444	2				
16	Langsa	4.480000	97.96333	495	485	304	24
9	410	-953	1				
17	Lhokseumawe	5.188056	97.14028	882	853	587	29
3	598	-1449	1				
18	Nagan Raya	4.124841	96.49298	384	373	180	12
3	284	-576	2				

19		Pidie	5.074266	95.94097	1151	1144	718	58
7	799	-2097		1				
20		Pidie Jaya	5.154806	96.19513	168	161	85	6
5	199	-342		1				
21		Sabang	5.894693	95.31930	31	28	9	1
8	30	-54		1				
22		Simeulue	2.583333	96.08333	213	201	80	14
9	124	-341		1				
23		Subulussalam	2.644993	98.01652	415	371	202	7
5	174	-407		1				

```

> #Uji validitas dengan Indeks Silhouette
> # Filter data yang bukan noise (biasanya cluster 0 atau -1 adalah
noise)
> data_clustered <- tb_data[tb_data$st_dbscan_cluster > 0, ]
> cluster_labels <- data_clustered$st_dbscan_cluster
> # Hitung jarak antar titik (misalnya berdasarkan latitude dan long
itude)
> distance_matrix <- dist(data_clustered[, c("Latitude", "Longitude"
)])
> # Hitung silhouette
> silhouette_result <- silhouette(cluster_labels, distance_matrix)
> # Plot silhouette
> plot(silhouette_result, main = "Silhouette Plot ST-DBSCAN", border
= NA)
> # Rata-rata nilai silhouette
> mean_silhouette <- mean(silhouette_result[, 3])
> cat("Nilai rata-rata Silhouette Index:", mean_silhouette, "\n")
Nilai rata-rata Silhouette Index: -0.1431668
> #OPTICS
> library(ggplot2)
> library(dbSCAN)
> library(factoextra)
> library(RColorBrewer)

```

Warning message:

```

package 'RColorBrewer' was built under R version 4.3.1
> # Terapkan OPTICS dengan parameter yang sesuai
> coordinates_st <- tb_data %>% select(Latitude, Longitude)
> optics_model <- optics(coordinates_st, minPts = 5, eps = Inf)
> # Ekstrak cluster menggunakan metode 'xi'
> optics_clusters <- extractXi(optics_model, xi = 0.03)
> # Tambahkan hasil cluster ke dataframe
> tb_data$optics_cluster <- optics_clusters$cluster
> tb_data$optics_cluster
 [1] 1 2 1 1 2 2 2 2 2 1 2 1 2 2 2 2 2 1 1 1 1 0 2
attr(,"class")
 [1] "numeric" "xics"
> # Tentukan warna lebih menonjol menggunakan palet dari RColorBrewer
> colors <- brewer.pal(n = max(tb_data$optics_cluster) + 1, "Set1")
# Palet warna cerah
> # Visualisasi hasil clustering dengan warna lebih menonjol

```

```

> ggplot(tb_data, aes(x = Longitude, y = Latitude, color = factor(op
tics_cluster))) +
+   geom_point(size = 3) +
+   scale_color_manual(values = colors) + # Menggunakan warna lebih
menonjol
+   labs(title = "Clustering Spatio-Temporal Kasus TB di Kabupaten A
ceh (2019-2023) - Metode OPTICS",
+         x = "Longitude", y = "Latitude", color = "Cluster ID") +
+   theme_minimal() +
+   theme(plot.title = element_text(hjust = 0.5, face = "bold"), #
Membuat judul lebih menonjol
+         legend.position = "right") # Menempatkan legenda di sebel
ah kanan
> # Menampilkan hasil cluster dalam bentuk tabel
> print(tb_data)

```

Kabupaten	TB_2019	TB_2020	TB_2021	TB_2022	TB_2023	Latitude	Longitude	Delta_TB	st_dbscan_cluster
1 Aceh Barat	311	259	184	290	291	4.454275	96.15270	-753	1
2 Aceh Barat Daya	202	232	247	262	315	3.051264	97.33680	-628	1
3 Aceh Besar	408	347	355	437	439	5.452917	95.47778	-1108	1
4 Aceh Jaya	212	155	120	143	148	4.787368	95.64580	-482	1
5 Aceh Selatan	374	281	415	747	793	3.311506	97.35166	-1024	1
6 Aceh Singkil	208	95	67	196	197	2.358946	97.87216	-369	1
7 Aceh Tamiang	426	259	447	451	453	4.232887	98.00289	-1130	1
8 Aceh Tengah	180	189	147	266	272	4.448264	96.83510	-510	2
9 Aceh Tenggara	129	58	160	345	365	3.308867	97.69823	-327	1
10 Aceh Timur	567	578	648	813	821	5.255443	95.98855	-1785	1
11 Aceh Utara	901	762	671	1016	1024	4.978633	97.22214	-2326	1
12 Banda Aceh	997	379	604	1108	1156	5.550000	95.31667	-1932	1
13 Bener Meriah	67	100	62	104	115	4.774835	97.00684	-218	1
14 Bireuen	848	740	729	876	881	5.182540	96.89005	-2312	1
15 Gayo Lues	199	105	149	226	235	3.955165	97.35166	-444	2
16 Langsa	410	249	304	485	495	4.480000	97.96333	-953	1
17 Lhokseumawe	598	293	587	853	882	5.188056	97.14028	-1449	1

18	Nagan Raya	284	123	180	373	384	4.124841
96.49298	-576		2				
19	Pidie	799	587	718	1144	1151	5.074266
95.94097	-2097		1				
20	Pidie Jaya	199	65	85	161	168	5.154806
96.19513	-342		1				
21	Sabang	30	18	9	28	31	5.894693
95.31930	-54		1				
22	Simeulue	124	149	80	201	213	2.583333
96.08333	-341		1				
23	Subulussalam	174	75	202	371	415	2.644993
98.01652	-407		1				

optics\_cluster

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

1  
2  
1  
1  
2  
2  
2  
2  
2  
1  
2  
1  
2  
2  
2  
2  
2  
1  
1  
1  
1  
0  
2

```

> # Filter data yang bukan noise (cluster != 0 atau -1 untuk noise)
> library(cluster)
> non_noise_optics <- tb_data[tb_data$optics_cluster != 0, ]
> non_noise_clusters_optics <- non_noise_optics$optics_cluster
> # cek apakah ada lebih dari 1 cluster untuk menghitung Silhouette
Score
> if (length(unique(non_noise_clusters_optics)) > 1) {
+   silhouette_scores_optics <- silhouette(non_noise_clusters_optics
+   , dist(non_noise_optics[, c("Latitude", "Longitude")]))
+
+   # Rata-rata silhouette score untuk OPTICS
+   avg_silhouette_optics <- mean(silhouette_scores_optics[, 3])
+   print(paste("Silhouette score rata-rata untuk OPTICS:", avg_silh
+   ouette_optics))
+

```

```

+ # Plot silhouette score untuk OPTICS
+ plot(silhouette_scores_optics, border = NA, main = "Silhouette Plot OPTICS")
+ } else {
+   print("Tidak bisa menghitung Silhouette Score untuk OPTICS karena hanya ada satu cluster.")
+ }
[1] "Silhouette Score rata-rata untuk OPTICS: 0.409496753394924"
> # Filter data yang bukan noise (cluster != 0 atau -1 untuk noise)
> non_noise_stdbscan <- tb_data[tb_data$st_dbscan_cluster != 0, ]
> non_noise_clusters_stdbscan <- non_noise_stdbscan$st_dbscan_cluster
> # Cek apakah ada lebih dari 1 cluster untuk menghitung Silhouette Score
> if (length(unique(non_noise_clusters_stdbscan)) > 1) {
+   silhouette_scores_stdbscan <- silhouette(non_noise_clusters_stdbscan, dist(non_noise_stdbscan[, c("Latitude", "Longitude"))])
+
+   # Rata-rata silhouette score untuk ST-DBSCAN
+   avg_silhouette_stdbscan <- mean(silhouette_scores_stdbscan[, 3])
+   print(paste("Silhouette Score rata-rata untuk ST-DBSCAN:", avg_silhouette_stdbscan))
+
+   # Plot silhouette score untuk ST-DBSCAN
+   plot(silhouette_scores_stdbscan, border = NA, main = "Silhouette Plot ST-DBSCAN")
+ } else {
+   print("Tidak bisa menghitung Silhouette Score untuk ST-DBSCAN karena hanya ada satu cluster.")
+ }
[1] "Silhouette Score rata-rata untuk ST-DBSCAN: -0.143166789077441"
> anova_stdbscan <- aov(Delta_TB ~ factor(st_dbscan_cluster), data = tb_data)
> # Menampilkan hasil ANOVA untuk ST-DBSCAN
> summary(anova_stdbscan)
              Df    Sum Sq Mean Sq F value Pr(>F)
factor(st_dbscan_cluster)  1    631086    631086    1.268  0.273
Residuals                21  10453613    497791
> # Jika hasil ANOVA signifikan (p-value < 0.05), lanjutkan dengan uji Tukey HSD
> if (summary(anova_stdbscan)[[1]][["Pr(>F)"]][1] < 0.05) {
+   tukey_stdbscan <- TukeyHSD(anova_stdbscan)
+   print(tukey_stdbscan)
+ }

```