

**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)**

TUGAS AKHIR



Nama : Anas Wahyu Sulistyio

Nim : 16611019

**PROGAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2020**

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul

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)

Nama Mahasiswa

: Anas Wahyu Sulistyو

NIM

: 16611019

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 5 Oktober 2020

Dosen Pembimbing 1

البعثة الإسلامية العالمية
الجامعات الإسلامية



Dr. techn. Rohmatul Fajriyah, S.Si., M.Si.

HALAMAN PENGESAHAN

TUGAS AKHIR

IMPLEMENTASI METODE *DENSITY-BASED SPATIAL OF APPLICATIONS WITH NOISE (DBSCAN)* DALAM MENGELOMPOKKAN PENYEBARAN TUBERKULOSIS
(Studi Kasus: Klaster Data Penyebaran Tuberkulosis *High Burden Country* (HBC) pada Tahun 2018)


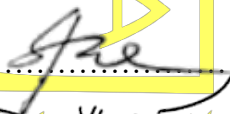

Nama Mahasiswa : Anas Wahyu Sulisty

NIM : 16611019

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL: 22 Oktober 2020**

Nama Penguji:

Tanda Tangan

1. Dr. techn. Rohmatul Fajriyah, S.Si., M.Si 
2. Saepudin, S.Si., M.Si., Ph.D., Apt. 
3. Achmad Fauzan, S.Pd., M.Si. 

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Prof. Riyanto, S.Pd., M.Si., Ph.D.

KATA PENGANTAR

Assalamu'alaikum Wr.Wb

Alhamdulillah Robbil 'Alamin, segala puji bagi Allah SWT Tuhan Semesta Alam. Karena atas berkat rahmat, hidayah, karunia dan nikmat-Nya penulis dapat menyusun dan menyelesaikan tugas akhir yang berjudul **“Implementasi Metode *Density-Based Spatial Clustering Of Applications With Noise (DBSCAN)* dalam Mengelompokkan Penyebaran Tuberkulosis”** dengan (Studi Kasus: Klaster Data Penyebaran Tuberkulosis *High Burden Country (HBC)* pada Tahun 2018 dengan lancar dan sebaik-baiknya sebagai salah satu persyaratan dalam menyelesaikan jenjang strata satu di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Shalawat serta salam tak lupa penulis haturkan kepada junjungan Nabi Agung, Nabi Muhammad SAW beserta keluarga, sahabat dan umatnya yang selalu membimbing ke jalan yang penuh berkah ini.

Penyusunan tugas akhir ini tidak terlepas dari banyak dukungan, bantuan, arahan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Kedua Orang Tua saya, Bapak (Subagiyo) dan Ibu (Suginem) yang telah banyak berkorban dan berjuang mulai dari tenaga, materi, pikiran untuk dapat menyekolahkan saya diperguruan tinggi, selalu memotivasi, menghibur, dan mendoakan yang terbaik sehingga saya selalu mendapatkan hidayah dari Allah SWT selama masa perkuliahan hingga saat ini.
2. Kedua kakak saya, Kiki dan Neira yang telah banyak membantu saya, sehingga saya dapat kuliah dan menyelesaikan kuliah saya ini.
3. Bapak Prof. Riyanto S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
4. Bapak Edy Widodo, S.Si., M.Si. Selaku Ketua Progam Studi Statistika, Universitas Islam Indonesia.

5. Ibu Dr. techn. Rohmatul Fajriyah, S.Si., M.Si. selaku Dosen Pembimbing Tugas akhir saya yang sudah meluangkan waktunya dalam membimbing penyusunan tugas akhir ini, dan selalu memberikan semangat dan motivasi dalam prosen penyusunan tugas akhir ini.
6. Bapak dan Ibu Dosen Program Studi Statistika, Universitas Islam Indonesia yang sudah membimbing dikelas dari semester awal hingga akhir, sehingga saya dapat memperoleh banyak pengalaman dan ilmu pengetahuan yang dapat saya tuangkan dalam penyusunan tugas akhir ini.
7. Sahabat seperjuangan bimbingan TA Ibu Ema, yang telah bekerjasama dengan baik, menemani penyusunan tugas akhir dan bertukar pikiran dalam penyempurnaan penyusunan skripsi hingga terlaksananya sidang pendadaran.
8. Teman-teman jurusan statistika angkatan 2016 yang sudah menemani, membantu dan banyak bertukar ilmu selama 4 tahun belajar bersama untuk meraih gelar sarjana ini.
9. Serta seluruh pihak-pihak lainnya yang tidak dapat disebutkan satu per satu, yang telah banyak membantu terselesaikannya tugas akhir ini dengan penuh drama dan perjuangan.

Meskipun tugas akhir ini telah disusun dengan sebaik-baiknya, namun penulis menyadari bahwa tugas akhir ini masih banyak membutuhkan bantuan koreksi dari berbagai pihak, baik berupa saran, kritik, pembaharuan, dan penyempurnaan. Penulis mengucapkan terima kasih dan mohon maaf atas segala kekurangan yang ada, karena kesempurnaan hanyalah milik Allah SWT. Semoga tugas akhir ini dapat bermanfaat tidak untuk penulis saja, namun juga dapat bermanfaat sebagai bahan penelitian yang dapat diaplikasikan dimasa mendatang yang berguna bagi banyak orang. Akhirul kata, wabillahi taufik walhidayah.

Wassalamualaikum Wr.Wb

Yogyakarta, 20 Oktober 2020

Penulis

(Anas Wahyu Sulisty)

DAFTAR ISI

HALAMAN Sampul	i
HALAMAN PERSETUJUAN PEMBIMBING	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN.....	x
PERNYATAAN.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terkait Penyakit Tuberkulosis.....	5
2.2 Penelitian Terkait DBSCAN	6
BAB III LANDASAN TEORI.....	11
3.1 Tuberkulosis	11
3.2 Analisis Deskriptif.....	11
3.3 Asumsi Analisis Kelompok.....	11
3.3.1 Sampel Representatif.....	12
3.3.2 Tidak Ada Multikolinieritas	12
3.3.3 Analisis Komponen Utama.....	13
3.4 <i>Data Mining</i>	17
3.5 Klaster.....	19
3.6 <i>Density-Based Spatial Clustering of Application with Noise</i>	21
BAB IV METODOLOGI PENELITIAN	24

4.1	Populasi dan Sampel Penelitian.....	24
4.2	Sumber dan Jenis Data Penelitian	24
4.3	Variabel Penelitian	24
4.4	Metode Pengumpulan Data	25
4.5	Metode Analisis Data	25
4.6	Langkah-langkah Penelitian	25
BAB V	HASIL DAN PEMBAHASAN	28
5.1.	Analisis Deskriptif.....	28
5.2.	Asumsi Cluster	31
5.2.1	Sampel Representatif.....	31
5.2.2	Uji Multikolinieritas	31
5.3.	Analisis Komponen Utama.....	32
5.4.	Analisis klaster dengan DBSCAN.....	33
5.5.	Intepretasi Hasil Klaster	35
BAB VI	PENUTUP	38
4.1	Kesimpulan.....	38
4.2	Saran	38
DAFTAR PUSTAKA	39
LAMPIRAN	42

DAFTAR TABEL

Tabel 2.1. Perbandingan Penelitian ini dengan Penelitian Sebelumnya	8
Tabel 4.1. Definisi Variabel Penelitian	24
Tabel 4.2. Keterangan <i>Flowchart</i>	27
Tabel 5.1. Statistika Deskriptif (dalam % populasi penduduk).....	28
Tabel 5.2. Matriks Korelasi	31
Tabel 5.3. Nilai <i>Eigenvalues</i> dan Proporsi	33
Tabel 5.4. Hasil Klaster DBSCAN dengan Epsilon 0,65 dan MintPts 2.....	35
Tabel 5.5. Profilisasi Hasil Klaster.....	36



DAFTAR GAMBAR

Gambar 3.1. Titik Border dan Titik Core	22
Gambar 3.2. Ilustrasi <i>Density-Reachable</i>	22
Gambar 3.3. Ilustrasi <i>Density-Connected</i>	23
Gambar 4.1. <i>Flowchart</i> Penelitian	26
Gambar 5.1. Grafik Statistika Deskriptif Variabel TB/HIV dan TB Paru	29
Gambar 5.2. Grafik Statistika Deskriptif Variabel EPTB dan Anak-anak	30
Gambar 5.3. Grafik Statistika Deskriptif Variabel Perempuan dan Laki-laki	30
Gambar 5.4. Grafik KNN displot	34
Gambar 5.5. Visualisasi hasil kluster	34



DAFTAR LAMPIRAN

Lampiran 1 <i>Script Program Software R</i>	43
Lampiran 2 Data Tuberkulosis menurut HBC tahun 2018 (dalam persentase populasi di setiap negara).....	45



PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 3 November 2020

A 6000 Rupiah Indonesian postage stamp is placed over the signature. The stamp features the Garuda Pancasila emblem and the text "POS TERAI EMPER", "1890AHF700644663", and "6000 RUPIAH".

(Faisal Ardiansyah)



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)

Anas Wahyu Sulisty
Program Studi Statistika, Fakultas MIPA
Universitas Islam Indonesia

INTISARI

Tuberkulosis merupakan salah satu penyakit menular yang disebabkan oleh bakteri Mycobacterium tuberculosis. Menurut World Health Organization (WHO) Sebagian besar kasus TB pada 2018 berada di wilayah WHO Asia Tenggara 44%, Afrika 24%, Pasifik Barat 18%, Mediterania Timur 8%, Amerika 3% dan Eropa 3%. Terdapat delapan negara menyumbang dua pertiga dari total global: India 27%, Cina 9%, Indonesia 8%, Filipina 6%, Pakistan 6%, Nigeria 4%, Bangladesh 4% dan Afrika Selatan 3%. Tujuan dari penelitian ini untuk mengetahui hasil analisis deskriptif penyebaran penyakit TBC berdasarkan data High Burden Country (HBC) dan mengklasterkan penyebaran penyakit TBC menggunakan metode Density-based Spatial Clustering with Noise (DBSCAN). Hasil dari analisis yang dilakukan, didapatkan hasil klaster sebanyak 3, dimana pada klaster 1 terdapat 22 negara dengan beban penyakit tuberkulosis rendah, pada klaster 2 sebanyak 2 negara dengan beban penyakit tuberkulosis tinggi dan pada klaster 3 sebanyak 2 negara dengan beban penyakit tuberkulosis sedang, sedangkan sisanya sebanyak 1 negara dianggap sebagai noise yaitu Negara India.

Kata Kunci: Tuberkulosis, *Data Mining*, Analisis Klaster, DBSCAN.

IMPLEMENTATION OF DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE (DBSCAN) METHOD IN CLUSTER OF TUBERCULOSIS

(Case Study: High Burden Country (HBC) Tuberculosis Data Cluster in 2018)

Anas Wahyu Sulistyو

Department of Statistics, Faculty of Mathematics and Natural Sciences

Islamic University of Indonesia

ABSTRACT

Tuberculosis is an infectious disease caused by the bacteria Mycobacterium tuberculosis. According to the World Health Organization (WHO), most TB cases in 2018 were in the WHO region, Southeast Asia 44%, Africa 24%, the West Pacific 18%, the East Mediterranean 8%, America 3% and Europe 3%. Eight countries account for two thirds of the global total: India 27%, China 9%, Indonesia 8%, Philippines 6%, Pakistan 6%, Nigeria 4%, Bangladesh 4% and South Africa 3%. The purpose of this study was to determine the results of a descriptive analysis of the spread of TB disease based on High Burden Country (HBC) data and to cluster the spread of TB using the Density-based Spatial Clustering with Noise (DBSCAN) method. The results of the analysis conducted showed that there were 3 clusters, where in cluster 1 there were 22 countries with a low burden of tuberculosis, in cluster 2 there were 2 countries with a high burden of tuberculosis and in cluster 3 there were 2 countries with moderate burden of tuberculosis, while the remaining 1 country is considered as noise, namely India.

Keywords: *Tuberculosis, Tuberkulosis, Data Minin,, Clustering, DBSCAN.*

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Tuberkulosis merupakan salah satu penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Terdapat beberapa jenis *Mycobacterium* diantaranya adalah *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. Leprae* dan masih banyak lagi. Selain bakteri *Mycobacterium tuberculosis* yang dapat menyebabkan gangguan pada saluran pernafasan juga diketahui sebagai MOTT (*Mycobacterium Other Than Tuberculosis*) yang dapat mengganggu dalam pengobatan TBC (Misnadiarly, 2006).

Tuberkulosis menjadi salah satu permasalahan kesehatan di dunia karena penderita penyakit ini terus bertambah baik di negara maju maupun di negara berkembang seperti di Indonesia. Menurut *World Health Organization* (WHO) “sebagian besar kasus TB pada 2018 berada di wilayah WHO Asia Tenggara 44%, Afrika 24%, Pasifik Barat 18%, Mediterania Timur 8%, Amerika 3% dan Eropa 3%. Delapan negara menyumbang dua pertiga dari total global: India 27%, Cina 9%, Indonesia 8%, Filipina 6%, Pakistan 6%, Nigeria 4%, Bangladesh 4% dan Afrika Selatan 3% juga terdapat 22 negara lain dalam daftar WHO dengan beban TB tinggi yang menyumbang 87% dari kasus dunia.” Berdasarkan hasil survei prevalensi Tuberkulosis pada laki laki 3 lebih besar dibandingkan dengan perempuan. Faktor utama yang menyebabkan bertambahnya kasus TBC adalah akibat dari kurang gizi, diabetes, merokok dan mengkonsumsi minuman beralkohol (Indah, 2018).

“Analisis *clustering* adalah salah satu cara pengelompokkan *record* pada suatu basis data berdasarkan kriteria tertentu. Hasil dari *clustering* digunakan untuk memberikan gambaran tentang apa yang terjadi pada basis data” (Han & Micheline, 2006). “*Clustering* melakukan pengelompokkan data tanpa berdasarkan kelas data tertentu. *Clustering* juga dapat dipgunakan untuk

memberikan label pada kelas data yang belum diketahui. Oleh karena itu *clustering* biasanya digolongkan sebagai metode *unsupervised learning*” (Witten dkk, 2011).

Metode *density-based spatial cluster of application with noise* (DBSCAN) adalah salah satu metode *clustering*. “Metode DBSCAN bekerja dengan menumbuhkan area-area dengan kepadatan yang cukup tinggi ke dalam kluster-kluster dan menemukan kluster-kluster dalam bentuk acak dalam suatu *database spatial* yang memuat suatu *noise*. DBSCAN mendefinisikan kluster sebagai himpunan maksimum dari titik-titik kepadatan yang tersambung (*densityconnected*). Semua objek yang tidak masuk ke dalam kluster manapun akan dianggap sebagai *outlier* atau *noise*” (Manning dkk, 2008).

Metode *density-based spatial cluster of application with noise* (DBSCAN) merupakan salah satu metode *cluster* mengacu pada densitas atau kepadatan. Kepadatan yang dimaksudkan yaitu dalam metode DBSCAN mengelompokkan wilayah dengan jarak yang telah ditentukan menggunakan nilai parameter Epsilon dan MinPts, sehingga dihasilkan suatu kelompok yang padat dengan jarak antar anggota kelompok yang beragam. Parameter Epsilon merupakan jarak maksimal antar titik pusat dengan titik anggota dalam suatu cluster. Sedangkan MinPts merupakan minimal anggota yang harus terpenuhi dalam sebuah kluster. Apabila kedua parameter tersebut telah terpenuhi, maka akan terbentuklah suatu kluster (Ester, dkk, 1996).

Berdasarkan permasalahan tersebut, peneliti akan menerapkan salah satu metode cluster non hierarki berdasarkan density (kepadatan) yaitu menggunakan metode *density-based spatial cluster of application with noise* (DBSCAN). Maka, penulis mengambil judul penelitian ini yaitu “**Implementasi Metode *Density-Based Spatial Clustering Of Applications With Noise* (DBSCAN) dalam Mengelompokkan Penyebaran Tuberkulosis**” dengan studi Kasus: Kluster Data Penyebaran Tuberkulosis *High Burden Country* (HBC) pada Tahun 2018.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka penulis menarik beberapa rumusan masalah yang akan diselesaikan pada penelitian ini yaitu:

- a. Bagaimana analisis deskriptif penyebaran penyakit TBC berdasarkan *High Burden Country* (HBC) pada Tahun 2018?
- b. Bagaimana hasil pengklasteran penyebaran penyakit TBC menggunakan metode *Density-based Spatial Clustering with Noise* (DBSCAN)?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini antara lain:

- a. Data yang digunakan dalam penelitian ini adalah data dari *World Health Organization* (WHO) *Global Tuberculosis Report* tahun 2018 dengan jumlah variable sebanyak 6.
- b. *Software* yang digunakan dalam penelitian ini adalah aplikasi *R Studio* 3.6.2 *software Microsoft Excel*.
- c. Metode klaster yang digunakan yaitu *Density-based Spatial Clustering with Noise* (DBSCAN).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini antara lain:

- a. Mengetahui hasil analisis deskriptif penyebaran penyakit TBC berdasarkan *High Burden Country* (HBC).
- b. Mengetahui hasil pengklasteran penyebaran penyakit TBC menggunakan metode *Density-based Spatial Clustering with Noise* (DBSCAN).

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat diantaranya adalah:

- a. Memudahkan penelitian selanjutnya dalam penentuan parameter *Epsilon* yang optimal.

- b. Mengetahui implementasi hasil pengklasteran dengan menggunakan metode *Density-based Spatial Clustering with Noise* (DBSCAN).
- c. Menjadi referensi bagi penelitian selanjutnya mengenai klasterr dengan data spasial.



BAB II

TINJAUAN PUSTAKA

Pada bagian ini akan dipaparkan beberapa penelitian sebelumnya yang mempunyai keterkaitan terhadap beberapa metode dan acuan yang digunakan oleh peneliti. Penelitian terdahulu digunakan sebagai kajian yang bertujuan mengetahui keterkaitan penelitian terdahulu dengan penelitian yang akan dilakukan untuk mengurangi kesalahan dan menghindari terjadinya duplikasi. Berikut beberapa penelitian dan publikasi terkait penyakit Tuberkulosis dan metode *Density-based Spatial Clustering with Noise*.

2.1 Penelitian Terkait Penyakit Tuberkulosis

(Heo dkk, 2005) Melakukan Penelitian yang berjudul “Pulmonary Tuberculosis: Another Disease Showing Klusters of Small Nodules.” Tujuan dari penelitian ini adalah untuk meninjau kasus tuberkulosis paru yang menunjukkan fitur CT tidak dapat dibedakan dari tanda *Sarcoid Galaxy*. Hasil dari penelitian ini menunjukkan pada radiografi awal terdapat 11 kelompok. Kemudian pada CT scan terdapat 19 kelompok.

(Wardani dkk, 2010) melakukan penelitian yang berjudul “Pentingnya Analisis *Klaster* Berbasis Spatial dalam Penanggulangan Tuberkulosis di Indonesia.” Penelitian ini dilakukan bertujuan untuk mengetahui pengelompokan penderita tuberkulosis menurut lokasi geografis dan dapat memberikan informasi tentang lokasi yang beresiko terhadap penyakit tersebut. Dari hasil penelitian ini didapatkan bahwa hasil *clustering* menunjukkan wilayah yang beresiko tinggi terhadap penyakit tuberkulosis adalah wilayah dengan determinan social yang rendah.

(Putra, 2012) melakukan penelitian yang berjudul “*Mycobacterium Tuberculosis* yang Resisten Terhadap Antibiotik Lini Pertama pada Pasien Tuberkulosis di RSUP Fatmawati.” Penelitian ini bertujuan untuk melakukan studi kasus *Mycobacterium Tuberculosis* yang Resisten Terhadap Antibiotik Lini

Pertama pada Pasien Tuberkulosis di RSUP Fatmawati dalam kurun waktu 3 tahun berdasarkan hasil pemeriksaan sputum BTA Laboratorium rujukan. Dari data resistensi terdapat 2 sampel yang ditemukan. Pertama adalah Rifampisin dengan nilai resistensi sebesar 35.2%. kedua adalah Streptomisin dengan nilai resistensi sebesar 19.3%.

(Sarwani & Sri Nurlaela, 2012) melakukan penelitian yang berjudul “Faktor Risiko *Multidrug Resistant Tuberculosis* (MDR-TB).” Penelitian ini dilakukan bertujuan untuk mengetahui faktor-faktor risiko MDR-TB. Dari hasil penelitian ini didapatkan bahwa faktor risiko yang berpengaruh pada kejadian MDR-TB yaitu motivasi rendah dan ketidakteraturan dalam minum obat. Adanya motivasi yang kuat dari penderita akan menyebabkan ketergantungan dalam minum obat. Salah satu masalah utama dalam penanganan kasus TB adalah dengan memotivasi pengidap agar mau menyelesaikan pengobatan sesuai dengan waktu yang telah ditentukan. Hal ini dapat dipengaruhi dengan kurangnya pengetahuan penderita tentang penyakitnya dan bagaimana mengobatinya, pelayanan yang kurang optimal dari penyelenggara fasilitas kesehatan, faktor budaya dan lain-lain.

(Rohman & Utami, 2019) melakukan penelitian yang berjudul “Pengelompokan Kasus Tuberkulosis Paru dan Upaya Promosi Kesehatan di Wilayah Mlati II Sleman Yogyakarta.” Penelitian ini bertujuan mengidentifikasi persebaran kasus tuberkulosis paru melalui pemetaan untuk upaya promosi kesehatan. Penelitian ini menggunakan sampel sebanyak 43 orang dengan menggunakan metode analisis *kernel density*. Hasil yang diperoleh dari penelitian ini menunjukkan bahwa berdasarkan jenis kelamin di Puskesmas Mlati II terdapat penderita laki-laki sebesar 67%. Kemudian berdasarkan data kelompok usia 17-25, 36-45, 56-65 dengan jumlah masing-masing 8 pasien, kelompok usia ini termasuk masa produktif dan lansia yang rentan terkena penyakit tersebut.

2.2 Penelitian Terkait DBSCAN

(Devi dkk, 2015) melakukan penelitian yang berjudul “Implementasi Metode *Clustering* DBSCAN pada Proses Pengambilan Keputusan.” Penelitian

ini mengacu pada pengaplikasian Metode DBSCAN dalam suatu langkah pengambilan keputusan yang berguna untuk membantu perusahaan dalam menentukan konsumen potensialnya. Penelitian ini didapatkan hasil sejumlah *klaster* dengan rata-rata nilai indeks validitas menggunakan Algoritma Silhouette lebih besar dari 0 dan mendekati 1, yang mana interval nilai indeks validitas algoritma ini yaitu -1 sampai dengan 1. Hal ini menandakan bahwa proses *clustering* menggunakan algoritma DBSCAN telah dapat dikategorikan baik.

(Silitonga, 2016) melakukan penelitian yang berjudul “Analisis Pola Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Dengan Menggunakan Metode *Dbscan Clustering*.” Penelitian yang dilakukan bertujuan untuk mendapatkan kelompok penyakit yang ada di sekelompok masyarakat yang berobat ke RSUP Haji Adam Malik Medan menggunakan metode DBSCAN. Hasil dari hasil analisis clustering yang dilakukan terhadap data pasien penggunaan BPJS Kesehatan di RSUP Haji Adam Malik Medan dihasilkan kesimpulan bahwa proses clustering dengan menggunakan algoritma DBSCAN Clustering menghasilkan 4 klaster yang menyimpan data-data penyakit pasien dengan karakteristik yang berdekatan.

(Saputera, 2019) melakukan penelitian yang berjudul “Implementasi *Clustering* pada Algoritma DBSCAN untuk Deteksi Kejadian Bencana Alam Pada Microblogging Twitter.” Penelitian ini bertujuan untuk menambahkan klasifikasi pada tahap *Clustering* pada algoritma DBSCAN dan membandingkan hasil *clustering* antara data yang menggunakan klasifikasi dan data yang tidak menggunakan klasifikasi untuk melihat peningkatan hasil *text mining* dari hasil *tweet* yang telah diolah. Hasil yang diperoleh dari penelitian ini menunjukkan nilai keseluruhan atau *overall similarity value* dari *klaster tweet* yang menggunakan klasifikasi pada tahap *clustering* lebih baik daripada data yang tidak menggunakan klasifikasi dan penggunaan klasifikasi terbukti meningkatkan kualitas daftar kata, yang ditunjukkan dengan *overall similarity value klaster* yang semakin mendekati *value* 1.

(Chanafi dkk, 2019) melakukan penelitian yang berjudul “Implementasi Algoritma *Clustering* Untuk Pengelompokan Pelanggan Retail Berdasarkan Skor

Recency, Frequency, Dan Monetary.” Penelitian yang dilakukan bertujuan untuk menyelidiki masalah pemetaan pelanggan yang berkaitan dengan *customer relationship management* dan juga penerapan dari metode DBSCAN dan *Fuzzy C-Means*. Hasil dari hasil analisis clustering yang dilakukan menghasilkan kesimpulan bahwa metode DBSCAN memiliki tingkat validitas kluster (SI total) yang lebih tinggi sebesar 0,918184744127465, sedangkan *Fuzzy C-Means* memiliki tingkat validitas kluster (SI total) yang lebih rendah yaitu 0,387801709898368.

(Kristanto dkk, 2020) melakukan penelitian yang berjudul *“Implementation DBSCAN Algorithm to Clustering Satellite Surface Temperature Data in Indonesia.”* Penelitian yang dilakukan bertujuan untuk mengelompokkan wilayah yang memiliki temperatur tinggi. Pengelompokkan ini dilakukan untuk mitigasi wilayah wilayah yang rentan terhadap kebakaran hutan. Hasil dari hasil analisis clustering yang dilakukan menghasilkan kesimpulan bahwa kluster 0 terdapat wilayah yang sering terjadi peristiwa kebakaran lahan maupun hutan. Kemudian pada kluster 3 adalah wilayah yang memiliki temperature yang tinggi.

Tabel 2.1. Perbandingan Penelitian ini dengan Penelitian Sebelumnya

Nama peneliti	Judul Penelitian	Tahun	Isi
Heo, dkk	<i>“Pulmonary Tuberculosis: Another Disease Showing Klusters of Small Nodules”</i>	2005	Tujuan dari penelitian ini adalah untuk meninjau kasus tuberculosis paru yang menunjukkan fitur CT tidak dapat dibedakan dari tanda <i>Sarcoid Galaxy</i> .
Wardani, dkk	<i>“Pentingnya Analisis Klaster Berbasis Spatial dalam Penanggulangan Tuberkulosis di Indonesia”</i>	2010	Penelitian ini dilakukan bertujuan untuk mengetahui pengelompokan penderita tuberculosis menurut lokasi geografis dan dapat memberikan informasi tentang lokasi yang beresiko terhadap penyakit tersebut.
Putra	<i>“Mycobacterium</i>	2012	Penelitian ini bertujuan untuk melakukan studi kasus

	<i>Tuberculosis yang Resisten Terhadap Antibiotik Lini Pertama pada Pasien Tuberkulosis di RSUP Fatmawati</i>		<i>Mycobacterium Tuberculosis yang Resisten Terhadap Antibiotik Lini Pertama pada Pasien Tuberkulosis di RSUP Fatmawati dalam kurun waktu 3 tahun berdasarkan halis pemeriksaan sputum BTA Laboratorium rujukan.</i>
Sarwani & Sri Nurlela	“Faktor Risiko <i>Multidrug Resistant Tuberculosis</i> (MDR-TB)”	2012	Penelitian ini dilakukan bertujuan untuk mengetahui berbagai factor risiko MDR-TB.
Rohman, & Utami	“Pengelompokan Kasus Tuberkulosis Paru dan Upaya Promosi Kesehatan di Wilayah Mlati II Sleman Yogyakarta”	2019	Penelitian ini bertujuan mengidentifikasi persebaran kasus tuberkulosis paru melalui pemetaan untuk upaya promosi kesehatan. Penelitian ini menggunakan sampel sebanyak 43 orang dengan menggunakan metode analisis <i>kernel density</i> .
Devi	“Implementasi Metode <i>Clustering</i> DBSCAN pada Proses Pengambilan Keputusan”	2015	Penelitian ini mengacu pada pengaplikasian Metode DBSCAN dalam suatu langkah pengambilan keputusan yang berguna untuk membantu perusahaan dalam menentukan konsumen potensialnya.
Silitonga	“Analisis Pola Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Dengan Menggunakan Metode <i>Dbscan Clustering</i> ”	2016	Penelitian yang dilakukan bertujuan untuk mendapatkan kelompok penyakit yang ada di sekelompok masyarakat yang berobat ke RSUP Haji Adam Malik Medan menggunakan metode DBSCAN.
Saputera	“Implementasi	2019	Tujuan dari Penelitian ini adalah untuk menambahkan

	<i>Klustering pada Algoritma DBSCAN untuk Deteksi Kejadian Bencana Alam Pada Microblogging Twitter</i>		klasifikasi pada tahap <i>Clustering</i> pada algoritma DBSCAN dan membandingkan hasil <i>clustering</i> antara data yang menggunakan klasifikasi dan data yang tidak menggunakan klasifikasi untuk melihat peningkatan hasil <i>text mining</i> dari hasil <i>tweet</i> yang telah diolah.
Chanafi	“Implementasi Algoritma <i>Clustering</i> Untuk Pengelompokan Pelanggan Retail Berdasarkan Skor <i>Recency, Frequency, Dan Monetary</i> ”	2019	Penelitian yang dilakukan bertujuan untuk mengetahui masalah pemetaan konsumen yang berkaitan dengan <i>customer relationship management</i> dan juga penerapan dari metode DBSCAN dan <i>Fuzzy C-Means</i> .
Kristanto, dkk	“ <i>Implementation DBSCAN Algorithm to Clustering Satellite Surface Temperature Data in Indonesia</i> ”	2020	Penelitian yang dilakukan bertujuan untuk mengelompokkan wilayah yang memiliki temperatur tinggi. Pengelompokan ini dilakukan untuk mitigasi wilayah wilayah yang rentan terhadap kebakaran hutan. Metode yang digunakan dalam penelitian ini menggunakan DBSCAN.
Penelitian yang penulis lakukan			
Anas Wahyu Sulisty	Implementasi <i>Density-Based Spatial Clustering of Applications with Noise</i> (DBSCAN)	2020	Pembeda pada penelitian yang saya lakukan dengan penelitian-penelitian sebelumnya yaitu terletak pada objek data yang peneliti gunakan. Data yang digunakan dalam penelitian ini adalah data penyebaran penyakit tuberculosi pada negara-negara yang masuk dalam HBC pada tahun 2018.

BAB III

LANDASAN TEORI

3.1 Tuberkulosis

“Tuberkulosis adalah suatu infeksi menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*, *Mycobacterium bovis* atau *Mycobacterium africanum*. Di negara berkembang seperti Indonesia, anak-anak biasanya terinfeksi oleh mikobakterium lainnya yang menyebabkan Tuberkulosis. Bakteri ini disebut *Mycobacterium bovis*, yang bisa menyebar melalui susu yang belum disterilisasi” (Misnadiarly, 2006).

Salah satu penyakit kronis yang telah lama dikenal oleh masyarakat luas adalah penyakit tuberkulosis. Robbert Kock melakukan penelitian pada tahun 1882 yang dapat membuktikan bukti jika penyakit tuberkulosis merupakan suatu penyakit infeksi yang disebabkan oleh bakteri yang bernama *Mycobacterium tuberculosis*. Kemudian Villenim adalah orang yang dapat membuktikan bahwa penyakit Tuberkulosis adalah suatu penyakit menular. Penyakit Tuberkulosis biasanya menyerang pada organ tubuh paru-paru, tetapi juga memungkinkan menyerang pada organ lain seperti kelenjar getah bening (*nodus lymphaticus*) (Misnadiarly, 2006).

3.2 Analisis Deskriptif

Analisis deskriptif merupakan metode sederhana yang biasanya digunakan dalam menganalisis data.

3.3 Asumsi Analisis Kelompok

Terdapat dua asumsi menurut (Hair, dkk, 1998) dalam suatu analisis kelompok antara lain sampel dapat dikatakan representatif dan tidak terdapat suatu multikolinieritas atau korelasi antar variabel.

3.3.1 Sampel Representatif

Pengambilan sampel pada suatu penelitian harus bisa menggambarkan populasi atau juga disebut dengan representative terhadap populasi. Penggunaan sampel yang representatif akan memberikan hasil penelitian yang maksimal dan sesuai dengan kondisi populasi yang sebenarnya. Tetapi jika dalam suatu penelitian data yang digunakan menggunakan data populasi maka dapat disimpulkan bahwa asumsi representatif telah terpenuhi. (Hair, dkk, 1998).

3.3.2 Tidak Terdapat Masalah Multikolinieritas

Multikolinieritas merupakan sebuah kejadian dimana terjadi korelasi yang tinggi antara dua atau lebih variabel kelompok. Dalam analisis multivariat kendala multikolinieritas merupakan suatu permasalahan yang harus dihindari, karena akan berpengaruh terhadap hasil akhir, sehingga dapat mengganggu analisis. Tetapi dalam analisis kluster efeknya berbeda, yaitu variabel-variabel yang mengalami kendala multikolinieritas secara signifikan dibobot lebih besar. Oleh karena itu peneliti diharuskan untuk melakukan pengujian multikolinieritas untuk mengetahui variabel mana yang menimbulkan kendala multikolinieritas. (Hair, dkk., 1998).

Koefisien korelasi salah satu cara untuk menguji multikolinieritas. Koefisien korelasi merupakan suatu bilangan atau indeks yang digunakan untuk menghitung nilai keterikatan hubungan antar variabel. Perhitungan koefisien korelasi yang dapat digunakan untuk data dengan skala pengukuran interval dan rasio adalah koefisien korelasi *pearson*. (Hasan, 2002). Berikut ini merupakan persamaan koefisien korelasi *pearson*:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.1)$$

Dengan:

r_{xy} = koefisien korelasi antara variabel x dan variabel y

x_i, y_i = variabel bebas x dan y pada data ke- i

\bar{x}, \bar{y} = rata-rata data variabel x dan variabel y

Berikut adalah uji hipotesis yang digunakan membuktikan ada atau tidaknya masalah multikolinieritas dalam data berdasarkan koefisien korelasi dan *P-value*:

a. Hipotesis Penelitian

$$H_0: \rho_{xy} = 0$$

$$H_1: \rho_{xy} \neq 0$$

b. Tingkat Signifikansi

$$\alpha = 0.05$$

c. Daerah Kritis

Tolak H_0 jika: $r_{xy} > r_{tabel}$ atau $P\text{-value} < \alpha$

d. Statistik Uji

Nilai r_{xy}

e. Keputusan

Tolak H_0 atau Gagal Tolak H_0

f. Kesimpulan

Dapat ditarik suatu kesimpulan dari keputusan yang diambil

“Apabila data penelitian masih mengandung masalah multikolinieritas dapat dilakukan beberapa hal untuk mengatasi masalah multikolinieritas, yaitu dapat mengurangi variabel dengan jumlah yang sama pada setiap set atau dapat menggunakan salah satu *distance measures*, seperti *Mahalanobis distance*” (Hair, dkk, 1998). Selain itu, “multikolinieritas juga dapat diatasi dengan menggunakan analisis komponen utama yang digunakan untuk mereduksi variabel menjadi beberapa faktor sehingga tidak mengandung kendala multikolinieritas”. (Supranto, 2004).

3.3.3 Analisis Komponen Utama (AKU)

Analisis komponen utama merupakan analisis yang sering digunakan untuk meringkas data tanpa mengurangi informasi pada data. Faktor-faktor diekstraksi akan menghasilkan faktor pertama yang memberikan peranan terbesar terhadap seluruh varian dari seluruh variabel asli, faktor kedua yang memberikan

peranan terbesar kedua, faktor ketiga yang memberikan peranan terbesar ketiga, dan seterusnya sehingga proses pencarian faktor bisa dihentikan setelah sumbangan terhadap seluruh varian variabel mencapai nilai minimal sebesar 60%. Dimungkinkan pula untuk mengekstraksi faktor sehingga faktor tidak berkorelasi sesamanya seperti di dalam Analisis Komponen Utama atau *Principal Component Analysis* (PCA) (Supranto, 2004).

“Dalam analisis komponen utama jumlah varian data yang terkandung dalam semua variabel asli dipertimbangkan, analisis ini direkomendasikan apabila tujuan peneliti adalah menentukan banyaknya faktor yang diekstraksi hingga sedikit mungkin tetapi menyerap sebagian besar informasi yang terdapat pada semua variabel asli untuk analisis multivariat selanjutnya.” (Supranto, 2004).

Terdapat dua uji yang harus dilakukan melakukan analisis faktor, yaitu pengujian terhadap data dengan menggunakan *uji bartlett* dan KMO. Berikut penjelasan untuk kedua uji tersebut:

a. Uji *Bartlett*

“Uji barlett digunakan untuk mengetahui apakah variabel yang digunakan berkorelasi dengan variabel lainnya. Pengujian dilakukan dengan menggunakan statistik *chi-square*, apabila nilai *chi-square* lebih dari nilai *chi-square* pada tabel maka variabel mengandung korelasi.” (Usman dan Sobari, 2013). Berikut uji hipotesis untuk uji *bartlett*:

i. Hipotesis

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

ii. Tingkat Signifikansi

$$\alpha = 0.05$$

iii. Daerah Kritis

$$\text{Tolak } H_0 \text{ jika: } \chi_{hitung}^2 > \chi_{(\alpha; db)}^2 \text{ atau } P\text{-value} < \alpha$$

iv. Statistik Uji

$$\text{Nilai } \chi_{hitung}^2 \text{ dan } P\text{-value}$$

v. Keputusan

Tolak H_0 atau Gagal Tolak H_0

vi. Kesimpulan

Dapat ditarik suatu kesimpulan dari keputusan yang diambil

b. Uji *Kaiser Mayer Olkin* (KMO)

“KMO adalah indeks yang digunakan untuk meneliti ketepatan analisis faktor. Analisis faktor tepat untuk digunakan ketika nilai KMO berada diantara nilai 0,5 hingga 1, apabila kurang dari 0,5 analisis faktor dikatakan tidak tepat untuk digunakan.” (Supranto, 2004). Berikut persamaan dalam perhitungan nilai KMO:

$$KMO = \frac{\sum_i^n \sum_{j \neq i}^n r_{ij}^2}{\sum_i^n \sum_{j \neq i}^n r_{ij}^2 + \sum_i^n \sum_{j \neq i}^n a_{ij}^2} \quad (3.2)$$

Dengan:

$$a_{ij} = -\frac{v_{ij}}{\sqrt{v_{ii}v_{jj}}}$$

a_{ij} = koefisien korelasi parsial dari variabel i dan variabel j

r_{ij} = koefisien korelasi sederhana dari variabel i dan variabel j

v_{ij} = invers matriks korelasi dari variabel i dan variabel j

ij = 1, 2, 3, 4... n

Terdapat dua metode dalam penentuan banyaknya faktor yang terbentuk, yaitu dengan menggunakan nilai *eigenvalues* dan persentase varian. Berikut penjelasan masing-masing metode:

1. Penentuan berdasarkan *Eigenvalues*

“Suatu *eigenvalues* menentukan besarnya peranan dari faktor terhadap varian seluruh variabel asli. Dalam pendekatan ini, hanya faktor dengan nilai *eigenvalues* yang lebih dari satu yang akan dimasukkan dalam model. Sedangkan yang kurang dari satu tidak dimasukkan karena dianggap tidak lebih baik dari variabel asli, sebab variabel asli telah dibakukan yang berarti rata-ratanya nol dan variannya satu.” (Supranto, 2004).

2. Penentuan berdasarkan Persentase Varian

Banyaknya faktor yang diekstraksi ditentukan sedemikian rupa sehingga kumulatif persentase varian yang diekstraksi oleh faktor mencapai suatu level tertentu yang memuaskan. Menurut Supranto (2004), ekstraksi faktor akan dihentikan ketika persentase varian kumulatif mencapai minimal 60% dari seluruh varian variabel asli.

Rotasi factor perlu dilakukan jika dalam ekstraksi faktor belum menghasilkan komponen faktor utama yang jelas. Rotasi faktor ini bertujuan untuk memperoleh struktur faktor yang lebih sederhana. Terdapat beberapa metode rotasi faktor yang dapat digunakan:

- a. *Varimax Method*, merupakan metode rotasi orthogonal untuk memperkecil jumlah indikator yang masih terapat *factor loading* tinggi pada setiap faktor.
- b. *Quartimax Method*, merupakan metode rotasi untuk memperkecil jumlah faktor yang digunakan dalam menjelaskan indikator.
- c. *Equamax Method*, merupakan metode gabungan antara *varimax method* dan *quartimax method* (Widarjono, 2010).

Analisis factor yang digunakan pada penelitian bertujuan untuk mendapatkan variabel yang mengandung multikolinieritas, maka penghitungan nilai faktor harus dilakukan untuk membentuk variabel baru guna melakukan analisis selanjutnya, dalam hal ini analisis kelompok. Berikut persamaan untuk menghitung nilai faktor:

$$F_i = w_{i1}X_1 + w_{i2}X_2 + \dots + w_{ij}X_j \quad (3.3)$$

Dengan:

F_i = skor faktor untuk faktor ke- i

w_{ij} = *weight* atau *factor score coefficient* faktor ke- i dan variabel ke- j

X_j = variabel ke- j

i = banyak faktor; $i = 1, 2, 3, \dots, n$

j = banyak variabel; $j = 1, 2, 3, \dots, n$

3.4 *Data Mining*

Data mining merupakan proses mengidentifikasi pola yang menarik dari basis data yang besar. Data mining adalah bagian dari proses *Knowledge Discover in Database* (KDD) yang bertujuan untuk mendapatkan informasi dan pola yang bias digunakan dalam suatu data. Kumpulan proses KDD terdiri dari langkah-langkah sebagai berikut: pembersihan data (*data cleaning*), integasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), *data mining*, evaluasi pola (*pattern evaluation*) dan *knowledge presentation* (Fu, 1997).

Menurut (Bramer, 2013) secara garis besar proses KDD dapat dijelaskan sebagai berikut:

1. *Data Selection*

Sebelum tahapan penggalian informasi pada Knowledge Discovery in Database (KDD) dimulai perlu dilakukan pemilihan data dari sekumpulan data operasional. Data hasil seleksi atau pemilihan data yang digunakan untuk proses data mining akan disimpan dalam suatu berkas baru yang terpisah dari *database* operasional.

2. *Pre-processing / Cleaning*

Proses *cleaning* pada data *Knowledge Discovery in Database* (KDD) perlu dilakukan sebelum proses data mining dilakukan. Proses *cleaning* bertujuan untuk menghilangkan data yang terduplikasi, memeriksa data yang tidak konsisten dan juga memperbaiki kesalahan pada data, seperti kesalahan cetak. Kemudian dilakukan proses *enrichment* atau disebut juga proses “memperkaya” data yang sudah ada dengan data atau informasi lainnya yang relevan dan yang diperlukan dalam proses KDD.

3. *Transformation*

Coding merupakan suatu proses transformasi data yang telah terpilih pada proses sebelumnya, sehingga data tersebut sesuai untuk proses *data mining*. Pada proses *coding* ini merupakan suatu proses kreatif dan sangat bergantung pada pola informasi tertentu.

4. *Data mining*

Suatu proses dalam pencarian suatu informasi yang menarik di dalam data yang telah terpilih dengan menggunakan suatu metode tertentu adalah *Data mining*. Metode-metode atau algoritma pada *data mining* bermacam-macam. Pemilihan metode atau algoritma yang tepat bergantung pada tujuan dan juga proses KDD secara keseluruhan.

5. *Interpretation / evaluation*

Informasi yang didapatkan dari proses data mining akan ditampilkan dalam bentuk yang mudah untuk dipahami. Pada tahapan ini adalah bagian dari suatu proses KDD yang juga disebut sebagai *interpretation*. Pada tahapan ini juga mencakup pemeriksaan pola atau informasi yang ditemukan akan bertentangan atau tidaknya dengan fakta atau hipotesis yang telah ada sebelumnya.

Perkembangan Teknologi Informasi telah menghasilkan sejumlah data dan database yang cukup besar dalam berbagai bidang. Penelitian dalam database dan teknologi informasi telah menerbitkan berbagai pendekatan untuk menyimpan dan memanipulasi data yang akan digunakan dalam pengambilan suatu keputusan. *Data mining* juga disebut sebagai proses ekstraksi informasi data yang sangat besar yang berguna dan bermakna. Tujuan dari metode ini adalah untuk menemukan pola yang sebelumnya tidak diketahui. Setelah ditemukannya pola-pola ini, maka pola-pola tersebut dapat digunakan untuk membuat keputusan tertentu untuk pengembangan data. Menurut (Bharati, 2010) terdapat tiga langkah yang terlibat dalam proses data mining, yaitu:

1. Eksplorasi

Eksplorasi adalah proses membersihkan data dan mengubahnya menjadi bentuk lain, variabel penting dipilih kemudian sifat data berdasarkan masalah dapat ditentukan.

2. Identifikasi Pola

Pada tahapan indentifikasi pola ini bertujuan untuk menyempurnakan dan mendefinisikan variabel menjadi lebih spesifik, kemudian membentuk identifikasi pola. Identifikasi dan penentuan pola akan menentukan suatu prediksi terbaik.

3. Deployment

Tahapan yang terakhir setelah mendapatkan pola adalah menyimpulkan pola tersebut untuk mendapatkan hasil yang diinginkan.

3.5 Klaster

Clustering merupakan pengelompokan data objek ke dalam *class* yang memiliki kesamaan. Kesamaan (*similarity*) diantara objek-objek ditaksir menurut *Distance Measure*. Hasil dari *clustering* adalah *klaster*, yang memiliki sekumpulan objek dengan kesamaan tertentu dalam satu *clustering* dan berbeda dengan objek pada *klaster* lain. Aktivitas untuk melakukan *clustering* disebut *Clustering analysis* (Han dkk, 2001).

Clustering tidak sama dengan klasifikasi karena didalam clustering tidak terdapat variable target untuk hasil akhir pengelompokkannya. Clustering tidak dapat digunakan untuk melakukan peramalan atau memperkirakan suatu nilai dari sebuah variable target. Tetapi metode *clustering* biasanya digunakan untuk mengelompokkan suatu data yang memiliki kemiripan dari data tersebut. Algoritma dari metode *klaster* bertujuan untuk membuat suatu dataset kedalam beberapa kelompok sesuai dengan karakteristiknya (Larose & Larose, 2005).

Menurut (Witten dkk, 2011), terdapat beberapa macam metode clustering seperti berikut ini:

1. Metode Partisi

Metode partisi merupakan metode untuk membangkitkan k pada sekumpulan data, kemudian setiap partisi mempresentasikan klaster. Metode ini bisa dilakukan dengan cara menetapkan pusat klaster awal, kemudian dilakukan realokasi obyek berdasarkan kriteria tertentu sampai didapatkan pengelompokan paling optimal. Partisi yang terbentuk harus memenuhi syarat, yaitu setiap *klaster* harus berisi minimal satu objek dan setiap objek harus membentuk satu *klaster*.

2. Metode hirarki

Metode hierarki merupakan metode yang membentuk beberapa kriteria berdasarkan suatu kumpulan data dari sebuah dekomposisi hierarki. Dalam

metode hierarki terdapat 2 jenis pendekatan, yaitu:

- a. Agglomerative: merupakan pembentukan suatu *klaster* dimulai dengan titik-titik yang membentuk suatu klaster individu, kemudian dilakukan penggabungan setiap titik yang berpasangan sampai tersisa satu titik.
- b. Divisive: merupakan pembentukan suatu *klaster* dimulai dengan pengelompokan keseluruhan titik, sehingga akan membentuk sebuah *klaster* yang besar. Kemudian *klaster* besar tersebut dibagi sampai hanya tersisa satu titik pada sebuah klaster (k klaster).

Menurut (Intaran, 2012), terdapat beberapa jenis klaster lainnya seperti berikut:

1. *Well-separated clusters*

Serangkaian titik yang terdapat di dalam *klaster*, dimana setiap titik memiliki kesamaan dibandingkan dengan titik yang berada pada luar *klaster*.

2. *Center-based clusters*

Merupakan sekumpulan objek dalam suatu *klaster* yang mempunyai kesamaan akan lebih cenderung identik dengan titik pusat dari suatu *klaster*, dibandingkan dengan titik pusat pada *klaster* lain. Titik pusat suatu klaster ini juga sering disebut sebagai centroid.

3. *Contiguous clusters*

Sebuah klaster adalah suatu kumpulan titik atau point, dimana titik yang terdapat di dalam klaster lebih dekat atau lebih identik daripada titik yang berada di luar klaster.

4. *Density-based clusters*

Merupakan suatu klaster dengan wilayah berdasarkan kepadatan, yang kemudian akan dibagi antara kepadatan daerah rendah dan kepadatan daerah tinggi. Klaster ini biasanya digunakan pada saat terdapat dataset atau kelompok data yang tidak beraturan dan terdapat data outlier.

5. *Property or Conceptual*

Digunakan untuk mencari suatu klaster yang memiliki fungsi untuk membagikan beberapa property umum atau mewakili suatu konsep tertentu.

6. *Described by an Objective function*

Digunakan untuk mendapatkan kluster yang meminimalkan atau memaksimalkan fungsi objektif. Terdapat perhitungan pada setiap kemungkinan cara untuk membagi titik ke dalam suatu kluster, kemudian dilakukan evaluasi goodness untuk setiap potensi sebuah kluster menggunakan suatu fungsi tertentu atau NP Hard.

3.6 *Density-Based Spatial Clustering of Application with Noise*

Density-Based Spatial Clustering of Application with Noise (DBSCAN) merupakan sebuah metode clustering yang membangun area berdasarkan kepadatan yang terkoneksi (*density connected*). DBSCAN berfungsi untuk mengidentifikasi kelompok-kelompok dalam kumpulan data spasial yang besar dengan acuan kepadatan lokal dari elemen-elemen basis data. DBSCAN juga biasa digunakan untuk menentukan sebuah hasil berupa informasi dapat dikategorikan sebagai *noise* atau *outlier* (Han dkk, 2011)

Setiap objek dari suatu kluster harus terdapat sejumlah minimum data. Semua objek yang tidak masuk ke dalam kluster akan dianggap sebagai *outlier* atau *noise*. Dengan menggunakan distribusi kepadatan dari titik-titik pada sebuah database, metode DBSCAN dapat mengkategorikan titik-titik tersebut ke dalam kelompok-kelompok yang terpisah untuk menandakan karakteristik tertentu. Menurut (Bäcklund dkk, 2011) terdapat 6 defini dan 2 lemma dalam proses komputasi DBSCAN antara lain:

Definisi 1. *Epsilon-neighborhood* dari satu titik

$$N_{Eps(p)} = \{ q \in D \mid dist(p, q) < eps \} \quad (3.4)$$

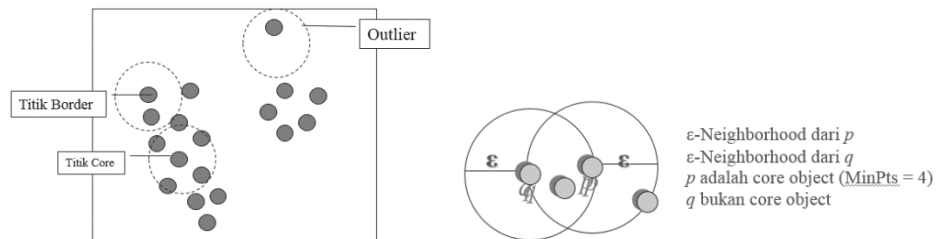
Sebuah titik yang masuk kedalam suatu kluster, minimal mempunyai satu titik lain yang letaknya lebih dekat ke titik tersebut dibandingkan dengan nilai epsilon.

Definisi 2. *Directly Density-Reachable*

$$p \in N_{Eps(q)} \quad (3.5)$$

Epsilon-neighborhood dari suatu *border point* cenderung memiliki titik yang lebih sedikit dibandingkan dengan *Epsilon-neighborhood* dari suatu *core*

point. Border point akan masih menjadi bagian dari suatu *klaster* apabila *border point* memiliki *Epsilon-neighborhood* dari suatu *core point* q sebagaimana yang ditunjukkan pada Gambar 2.



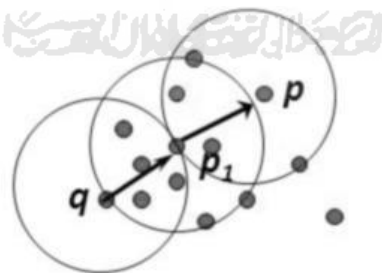
Gambar 3.1. Titik Border dan Titik Core

Agar titik q menjadi core point, titik tersebut perlu memiliki suatu jumlah minimum dari titik dalam Epsilon-neighborhood-nya. Kondisi core point ditunjukkan oleh persamaan 3

$$\| N_{Eps(Q)} \| \geq MinPts \quad (3.6)$$

Definisi 3. *Density-Reachable*

Suatu titik p merupakan density-reachable dari suatu titik q berdasarkan Epsilon dan MinPts jika terdapat suatu rantai titik-titik p_1, \dots, p_n , $p_1 = q$, $p_n = p$ dimana p_{i+1} merupakan directly density-reachable dari p_i . Pada Gambar 3 ditunjukkan ilustrasi dari suatu density-reachable.

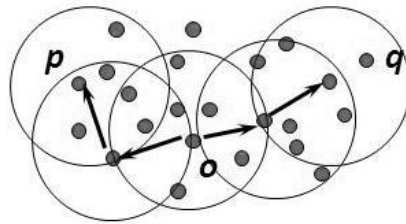


Gambar 3.2. Ilustrasi *Density-Reachable*

Definisi 4. *Density-Connected*

Terdapat kasus ketika dua border point akan masuk ke dalam suatu klaster yang sama tetapi dimana dua border point tersebut tidak membagi suatu core point spesifik. Dalam situasi ini titik-titik tersebut tidak akan menjadi density-reachable satu sama lainnya. Namun harus ada core point q yang mana menjadi density-reachable dari kedua border point tersebut. Pada Gambar 3 ditunjukkan

bagaimana density connectivity bekerja. Suatu titik p merupakan density-connected ke titik q berdasarkan Epsilon dan MinPts jika terdapat suatu titik o dimana keduanya p dan q merupakan density-reachable dari o berdasarkan Epsilon dan MinPts.



Gambar 3.3. Ilustrasi *Density-Connected*

Definisi 5. *Klaster*

Jika suatu titik p merupakan bagian dari suatu klaster C dan titik q merupakan density-reachable dari titik p berdasarkan jarak tertentu dan jumlah minimum titik-titik dalam jarak tersebut, maka q juga merupakan bagian dari klaster C . Dua titik dimiliki oleh klaster yang sama C , dimana p merupakan density-connected ke q berdasarkan jarak tertentu dan jumlah minimum titik-titik dalam jarak tersebut.

Definisi 6. *Noise*

Noise adalah sekumpulan titik-titik, dalam basis data, yang tidak masuk dalam klaster.

Lemma 1:

Suatu klaster dapat dibentuk dari salah satu core point dan akan selalu memiliki bentuk yang sama.

Lemma 2:

Jika p menjadi suatu core point dalam suatu klaster C dengan jarak minimum (Epsilon) dan suatu jumlah minimum titik dalam jarak tersebut (MinPts). Jika sekumpulan O merupakan density-reachable dari p berdasarkan Epsilon dan MinPts yang sama, maka C sama dengan sekumpulan O .

BAB IV

METODOLOGI PENELITIAN

4.1 Populasi dan Sampel Penelitian

Populasi yang digunakan dalam penelitian ini adalah data Penyebaran Tuberkulosis *High Burden Country* (HBC) menurut WHO pada tahun 2018. Sedangkan sampel yang digunakan dalam penelitian ini adalah menggunakan data populasi.

4.2 Sumber dan Jenis Data Penelitian

Data yang peneliti gunakan adalah data sekunder yang peneliti dapatkan dari website WHO dengan url <https://apps.who.int> yang diakses pada tanggal 20 September 2020.

4.3 Variabel Penelitian

Adapun variabel-variabel yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 4.1. Definisi Variabel Penelitian

Nama Variabel	Definisi Operasional Variabel
Negara	Negara merupakan suatu organisasi atau lembaga tertinggi dari kelompok masyarakat yang terdiri dari sekumpulan orang di wilayah tertentu, memiliki cita-cita untuk hidup bersama, serta memiliki sistem pemerintahan yang berdaulat.
TB HIV	TB HIV merupakan persentase pasien penderita penyakit tuberkulosis yang juga mengidap penyakit HIV pada setiap Negara
TB Paru	TB Paru merupakan persentase dari pasien yang terinfeksi bakteri <i>Mycobacterium tuberculosis</i> pada setiap negara. Namun, gangguan ini belum menyebar dari paru-paru ke bagian tubuh lainnya.

Nama Variabel	Definisi Operasional Variabel
EPTB	EPTB atau TB ekstra paru, merupakan persentase pasien yang terinfeksi bakteri <i>M. tuberculosis</i> telah menyebar ke jaringan dan organ tubuh selain paru-paru. Organ yang dapat terinfeksi bakteri penyebab TBC adalah kelenjar limpa, selaput otak, sendi, ginjal, tulang, kulit, bahkan alat kelamin.
Anak-anak	Persentase penderita Tuberkulosis pada anak-anak usia 0-14 tahun di setiap negara.
Perempuan	Persentase penderita Tuberkulosis berjenis kelamin perempuan di setiap negara.
Laki-laki	Persentase penderita Tuberkulosis berjenis kelamin laki-laki di setiap negara.

4.4 Metode Pengumpulan Data

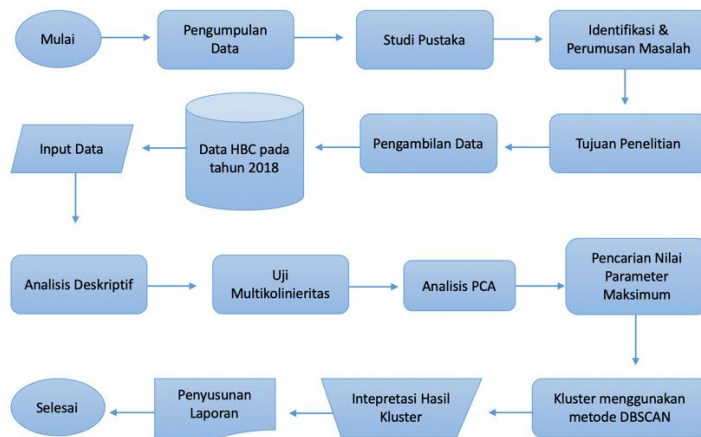
Data yang digunakan dalam penelitian ini merupakan data sekunder yang telah di download dari *Global Tuberculosis Report WHO* tahun 2019.

4.5 Metode Analisis Data

Metode analisis data yang digunakan pada penelitian ini adalah kluster dengan menggunakan metode *Density-based Spatial Clustering with Noise* (DBSCAN). Analisis ini digunakan untuk mengklusterkan Penyebaran Tuberkulosis *High Burden Country* (HBC) pada Tahun 2018. *Software* yang digunakan untuk penelitian ini adalah *software R studio* dengan versi 3.6.2.

4.6 Langkah-langkah Penelitian

Berikut ini adalah langkah-langkah atau tahapan yang dilakukan pada penelitian yang disajikan dalam bentuk *flowchart* berikut:





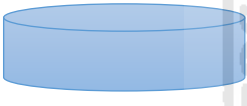

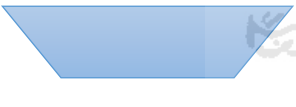


Gambar 4.1. *Flowchart* Penelitian

Keterangan:

1. Tahapan pertama yang dilakukan peneliti yaitu menentukan topik penelitian yang akan dilakukan, mencari studi pustaka atau referensi tentang penelitian-penelitian yang pernah dilakukan sebelumnya yang sesuai dengan topik yang telah ditentukan oleh peneliti kemudian mengidentifikasi rumusan masalah penelitian dan tujuan penelitian yang akan dilakukan.
2. Setelah tahapan pertama dilakukan, tahapan selanjutnya adalah pencarian data yang sesuai dengan topik penelitian yang akan dilakukan oleh peneliti. Pada penelitian ini peneliti mencari data pada website WHO.
3. Setelah mendapatkan data, tahapan selanjutnya adalah penginputan data pada *software* yang digunakan oleh peneliti dalam melakukan analisis.
4. Setelah melakukan penginputan data, tahapan selanjutnya adalah melakukan analisis deskriptif. pada penelitian ini menggunakan rata-rata (*mean*), nilai minimum dan nilai maksimum.
5. Tahapan selanjutnya adalah uji multikolinieritas untuk melihat apakah masih terdapat multikolinieritas antar variabel-variabel yang digunakan dalam penelitian ini.
6. Karena variabel yang digunakan dalam penelitian ini masih mengandung multikolinieritas, maka dilanjutkan pada analisis PCA untuk mengatasi masalah multikolinieritas tersebut.

7. Tahap selanjutnya adalah menganalisis data menggunakan metode kluster DBSCAN.
8. Setelah mendapatkan hasil *clustering*, selanjutnya adalah dilakukan interpretasi.
9. Hasil dari analisis yang telah dilakukan selanjutnya di proses kedalam laporan akhir penelitian.

Tabel 4.2. Keterangan *Flowchart*

Simbol	Keterangan
	Terminator simbol (Permulaan dari suatu kegiatan atau Pengakhiran)
	Simbol yang menyatakan proses <i>input</i> dan <i>output</i>
	<i>Magnetik Disk</i> (Hasil <i>Input</i> atau <i>Output</i> berupa Disk Magnetik)
	Proses Pengolahan Data atau Perhitungan Data
	<i>Manual Operation</i> (Pekerjaan atau Operasi Secara Manual)
	<i>Document</i> (Mencetak Hasil)
	Simbol Arus (Menghubungkan antara simbol yang satu dan lainnya)

BAB V

HASIL DAN PEMBAHASAN

Pada bab ini akan membahas tahapan-tahapan analisis serta hasil yang diperoleh untuk pengkalsteran penyebaran penyakit tuberkulosis tahun 2018. Analisis yang dilakukan menggunakan analisis deskriptif dan metode *Cluster DBSCAN*.

5.1. Analisis Deskriptif

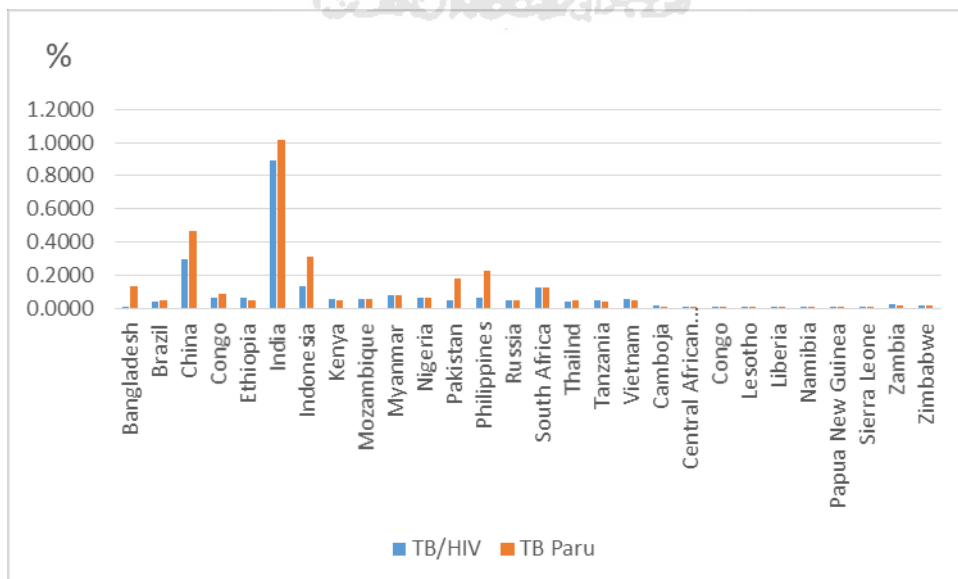
Analisis deskriptif adalah suatu analisis yang dilakukan untuk menilai karakteristik dari sebuah data atau gambaran umum tentang suatu data yang digunakan dalam penelitian, pada penelitian ini menggunakan rata-rata (*mean*), nilai minimum dan nilai maksimum.

Tabel 5.1. Statistika Deskriptif Penyebaran TB (dalam % populasi penduduk)

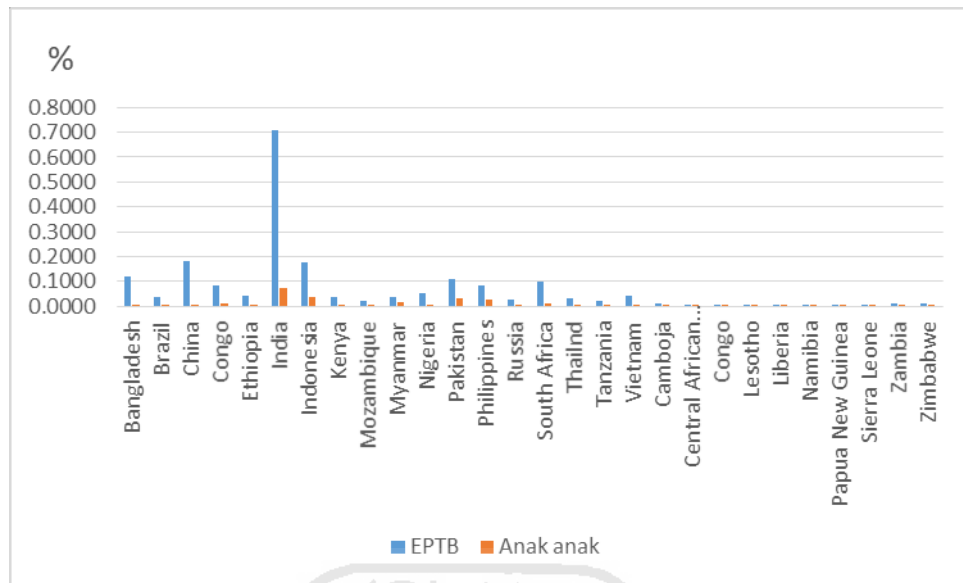
	TB / HIV	TB Paru	EPTB	Anak anak	Perempuan	Laki-laki
Mean	0,0805	0,1123	0,0701	0,0096	0,0451	0,0759
Minimum	0,0013	0,0032	0,0029	0,0002	0,0015	0,0023
Maximum	0,8917	1,0156	0,7060	0,0743	0,4211	0,7431

Berdasarkan **Tabel 5.1** diketahui bahwa rata-rata dari variabel TB/HIV memiliki nilai sebesar 0,0805% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Kemudian pada variabel TB Paru memiliki nilai rata-rata sebesar 0,1123% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Pada variabel ketiga yaitu EPTB memiliki nilai rata-rata sebesar 0,0701% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel penderita anak-anak memiliki nilai rata-rata sebesar 0,0096% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Pada variabel perempuan memiliki nilai rata-rata sebesar 0,0451% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel terakhir adalah laki-laki yang memiliki nilai rata-rata sebesar 0,0759% dari populasi penduduk negara-negara yang terdaftar dalam HBC.

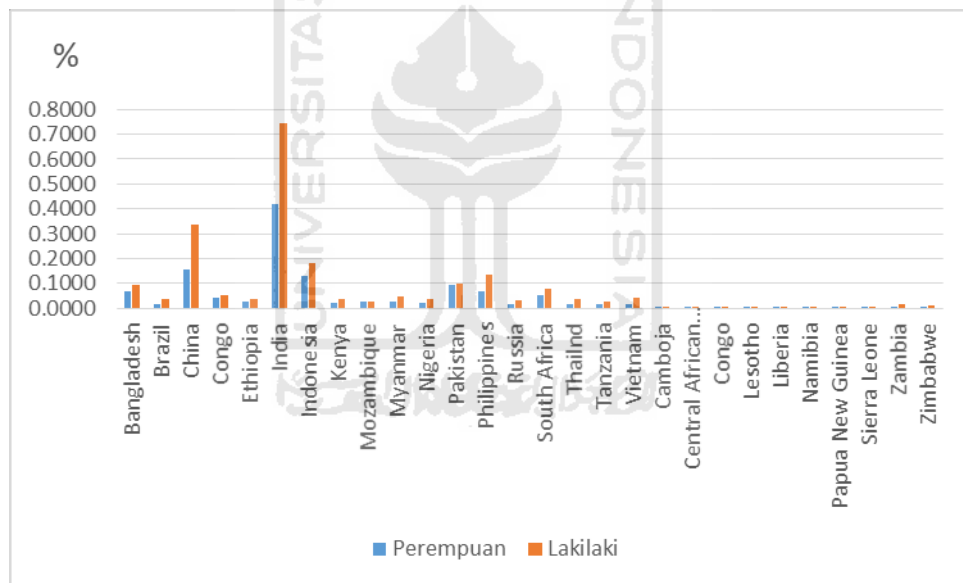
Pada nilai Minimum dari variabel TB/HIV memiliki nilai sebesar 0,0013% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel TB paru memiliki nilai sebesar 0,0032% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel EPTB sebanyak 0,0029% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel anak-anak memiliki nilai sebanyak 0,0002% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel perempuan memiliki nilai sebanyak 0,0015% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel laki-laki memiliki nilai sebanyak 0,0023% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Kemudian untuk nilai maksimum masing-masing variabel adalah pada variabel TB/HIV sebesar 0,8917% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel TB Paru sebesar 1,0156% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel EPTB sebesar 0,706% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel anak-anak sebesar 0,0743 dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel perempuan memiliki nilai sebanyak 0,4211% dari populasi penduduk negara-negara yang terdaftar dalam HBC. Variabel laki-laki memiliki nilai sebesar 0,7431% dari populasi penduduk negara-negara yang terdaftar dalam HBC.



Gambar 5.1. Grafik Statistika Deskriptif Variabel TB/HIV dan TB Paru



Gambar 5.2. Grafik Statistika Deskriptif Variabel EPTB dan Anak-anak



Gambar 5.3. Grafik Statistika Deskriptif Variabel Perempuan dan Laki-laki

Berdasarkan **Gambar 5.1** dapat dilihat bahwa negara India menempati urutan teratas berdasarkan TB/HIV, TB Paru, EPTB, Anak-anak, Perempuan dan Laki-laki. Menurut data WHO, negara India selalu menempati urutan pertama dalam masalah TBC dalam beberapa tahun terakhir.

5.2. Asumsi-asumsi Klaster

5.2.1 Sampel Representatif

Populasi pada penelitian ini adalah seluruh *High Burden Country* (HBC) Tuberkulosis menurut WHO. Penelitian ini menggunakan populasi sehingga asumsi sampel representatif sudah terpenuhi.

5.2.2 Uji Multikolinieritas

Uji multikolinieritas pada penelitian ini penulis menggunakan nilai korelasi dan dari hasil korelasi penulis hanya mengambil beberapa *output* saja. Berikut uji hipotesis dan hasil *output* nilai korelasi:

a. Hipotesis Penelitian

$$H_0: \rho_{xy} = 0$$

$$H_1: \rho_{xy} \neq 0$$

b. Tingkat Signifikansi

$$\alpha = 0,05$$

c. Daerah Kritis

Tolak H_0 jika: $P\text{-value} < \alpha$ atau nilai korelasi variabel $> r_{(0,05;28)}$

dimana $r_{(0,05;28)} = 0.374$

d. Statistik Uji

Tabel 5.2. Matriks Korelasi

	HIV	TB Paru	EPTB	Anak anak	Perempuan	Laki-laki
HIV	1	0.9594	0.9695	0.8165	0.9576	0.9724
TB Paru	0.9594	1	0.9750	0.8721	0.9924	0.9971
EPTB	0.9695	0.9750	1	0.8886	0.9896	0.9781
Anak-anak	0.8165	0.8721	0.8886	1	0.8999	0.8471
Perempuan	0.9573	0.9924	0.9896	0.8999	1	0.9891
Laki-laki	0.9724	0.9971	0.9781	0.8471	0.9891	1

e. Kesimpulan

Dengan tingkat signifikansi sebesar 5% data yang ada menjelaskan bahwa dari hasil statistik uji terdapat keputusan Tolak H_0 (terdapat multikolinieritas) dan Gagal Tolak H_0 (tidak ada multikolinieritas).

Dengan menggunakan uji multikolinieritas diatas dapat diartikan bahwa semua variabel yang digunakan terdapat kendala multikolinieritas. Sehingga dapat disimpulkan bahwa asumsi multikolinieritas pada penelitian ini tidak terpenuhi. Selanjutnya perlu dilakukan penanganan berupa reduksi factor terhadap variabel-variabel yang diteliti dengan menggunakan analisis komponen utama.

5.3. Analisis Komponen Utama

Sebelum melakukan analisis komponen utama perlu dilakukan uji *Bartlett* dan uji KMO.

1. Uji *Bartlett*

Uji *Bartlett* dilakukan untuk menguji hipotesis apakah antar variabel saling berkorelasi atau tidak berkorelasi. Berikut hasil uji *Bartlett*:

i. Hipotesis

$H_0: \rho = 0$ (Tidak terdapat Korelasi)

$H_1: \rho \neq 0$ (Terdapat Korelasi)

ii. Tingkat Signifikansi

$\alpha = 0,05$

iii. Daerah Kritis

Tolak H_0 jika $P\text{-value} < \alpha$

iv. Statistik Uji

Nilai $P\text{-value}$ yaitu $2,2e-16$

v. Keputusan

Tolak H_0 karena $P\text{-value} (2,2e-16) < \alpha (0,05)$

vi. Kesimpulan

Dengan tingkat signifikasni 5% data yang ada menjelaskan bahwa terdapat korelasi antar variabel.

2. KMO (*Kaiser Mayer Olkin*)

KMO bertujuan untuk menguji ketepatan analisis faktor, berdasarkan hasil *output Rstudio* diketahui bahwa nilai KMO yaitu sebesar 0,7582529. Nilai tersebut berada di antara nilai 0,5 sampai 1 yang menunjukkan bahwa analisis faktor bisa digunakan atau layak digunakan dalam penelitian ini.

Untuk menentukan banyaknya komponen yang terbentuk, peneliti menggunakan nilai *eigenvalues*. Komponen dengan nilai *eigenvalues* lebih dari satu akan dipertahankan, sedangkan apabila kurang dari satu maka faktor tidak diikutsertakan dalam model. Menurut Supranto (2004) “ekstraksi faktor dihentikan apabila persentase kumulatif varian paling sedikit 60% atau 75% dari seluruh variabel asli.”

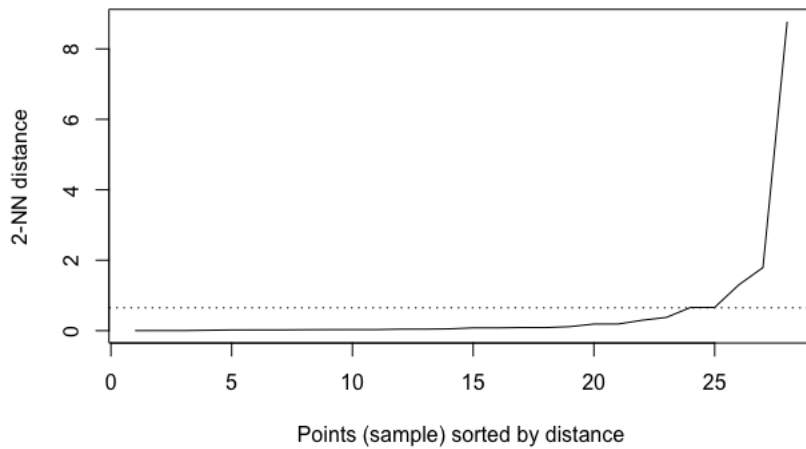
Tabel 5.3. Nilai *Eigenvalues* dan Proporsi

Faktor	<i>Eigenvalues</i>	<i>Proportion of Variance</i>	<i>Cumulative Proportion</i>
1	2.3889	0.9511	0.9511
2	0.47189	0.03711	0.98826
3	0.21743	0.00788	0.99614
4	0.14333	0.00342	0.99956
5	0.04702	0.00037	0.99993
6	0.02086	0.00007	1.00000

Berdasarkan **Tabel 5.3** dapat dilihat bahwa komponen 1 mempunyai nilai *Eigenvalues* lebih dari satu. Sehingga komponen yang digunakan pada analisis komponen utama ini sebanyak 1 komponen. Komponen 1 menunjukkan nilai varian sebesar 95.11% sehingga kumulatif varian yang dapat dijelaskan oleh komponen tersebut adalah sebesar 95.11% yang berarti lebih dari 60% yang artinya komponen tersebut dapat menjelaskan data.

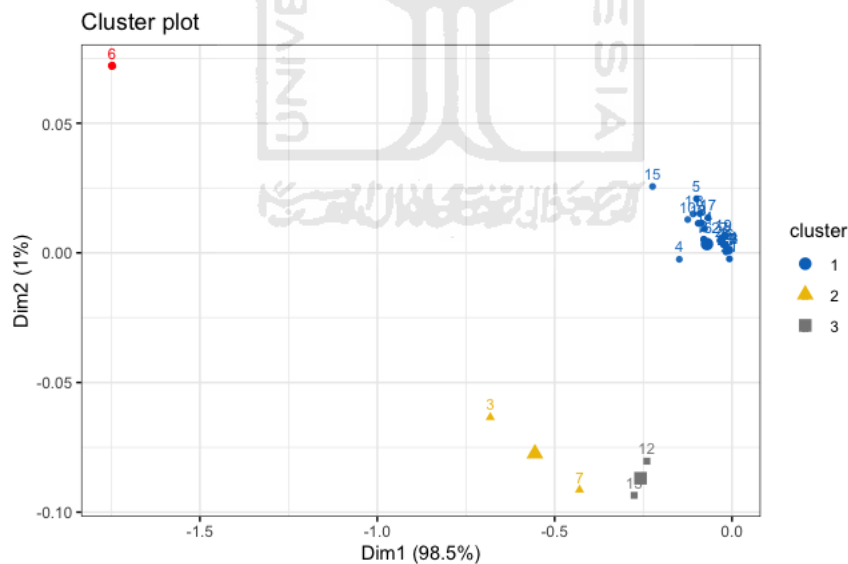
5.4. Analisis kluster dengan DBSCAN

Analisis yang akan digunakan dalam penelitian ini menggunakan kluster metode DBSCAN. Metode DBSCAN memerlukan parameter inputan yaitu parameter Epsilon dan MinPts. Adapun untuk menentukan nilai parameter yang akan digunakan dalam penelitian ini akan ditunjukkan pada grafik Gambar 5.4.



Gambar 5.4. Grafik KNN displot

Gambar 5.2 merupakan grafik KNN-displot yang digunakan untuk menentukan nilai epsilon yang akan digunakan dalam kluster DBSCAN. Pada grafik tersebut didapatkan nilai epsilon sebesar 15, kemudian setelah mendapatkan nilai epsilon akan dilanjutkan dalam analisis kluster DBSCAN. Pada penelitian ini, peneliti menggunakan nilai parameter MinPts sebesar 2.



Gambar 5.5. Visualisasi hasil kluster

Gambar 5.3 merupakan hasil visualisasi kluster DBSCAN dengan menggunakan nilai parameter Epsilon sebesar 0,65 dan nilai parameter MintPts sebesar 2. Pada hasil kluster ini menunjukkan, terdapat 3 kluster dimana pada kelompok kluster 1 sebanyak 23, kluster 2 sebanyak 2, kluster 3 sebanyak 2 dan

sisanya sebanyak 1 tidak masuk ke dalam klaster 1, klaster 2 dan klaster 3 atau juga disebut sebagai *noise*.

5.5. Interpretasi Hasil Klaster

Setelah melakukan uji asumsi klaster dan analisis komponen utama, selanjutnya akan diinterpretasikan hasil klaster metode DBSCAN dengan menggunakan nilai parameter Epsilon sebesar 0,65 dan MinPts sebesar 2.

Tabel 5.4. Hasil Klaster DBSCAN dengan Epsilon 0,65 dan MinPts 2

Klaster	Sampel
0	India
1	Bangladesh, Brazil, DRoCongo, Ethiopia, Kenya, Mozambique, Myanmar, Nigeria, Rusia, Afrika Selatan, Thailand, Tanzania, Vietnam, Kambija, Afrika Tengah, Congo, Lesotho, Liberia, Namibia, Papua New Ghuinea, Sierra Leone, Zambia, Zimbabwe
2	Pakistan, Filipina
3	China, Indonesia

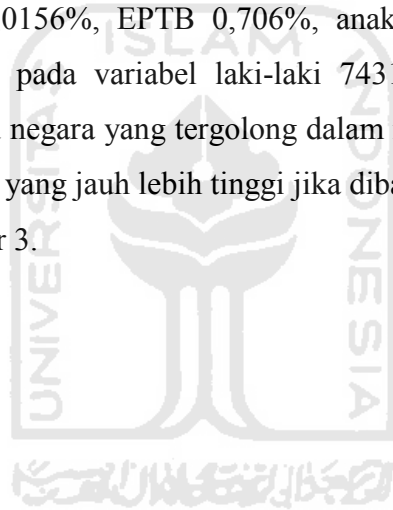
Berdasarkan tabel 5.4 terdapat dua klaster, dimana klaster 0 dianggap sebagai *noise*. Data negara-negara yang tergolong dalam noise adalah India. Kemudian klaster 1 merupakan klaster dengan tingkat penyebaran tuberkulosis rendah dimana terdapat sebanyak 23. Pada klaster 2 merupakan klaster dengan tingkat penyebaran tuberkulosis sedang dimana terdapat sebanyak 2 anggota antara lain adalah Pakistan dan Filipina. Klaster 3 merupakan klaster dengan tingkat penyebaran tuberkulosis tinggi dimana terdapat sebanyak 2 anggota antara lain adalah China dan Indonesia.

Tabel 5.5. Profilisasi Hasil Klaster

Cluster	Variabel	Statistik (dalam %)			Jumlah Anggota
		Rata-rata	Minimum	Maksimum	
Cluster 1	TB/HIV	0,0361	0,0013	0,1275	23
	TB Paru	0,0411	0,0032	0,1344	
	EPTB	0,0308	0,0029	0,1195	
	Anak-anak	0,0042	0,0002	0,0163	
	Perempuan	0,0172	0,0015	0,0680	
	Laki-laki	0,0274	0,0023	0,0913	
Cluster 2	TB/HIV	0,2130	0,1296	0,2964	2
	TB Paru	0,3887	0,3082	0,4692	
	EPTB	0,1789	0,1751	0,1828	
	Anak-anak	0,0217	0,0049	0,0385	
	Perempuan	0,1414	0,1296	0,1531	
	Laki-laki	0,2590	0,1821	0,3359	
Cluster 3	TB/HIV	0,0536	0,0448	0,0623	2
	TB Paru	0,2027	0,1791	0,2262	
	EPTB	0,0953	0,0831	0,1075	
	Anak-anak	0,0284	0,0277	0,0291	
	Perempuan	0,0816	0,0693	0,0940	
	Laki-laki	0,1173	0,1008	0,1339	
Noise	TB/HIV	0,8917	0,8917	0,8917	1
	TB Paru	1,0156	1,0156	1,0156	
	EPTB	0,7060	0,7060	0,7060	
	Anak-anak	0,0743	0,0743	0,0743	
	Perempuan	0,4211	0,4211	0,4211	
	Laki-laki	0,7431	0,7431	0,7431	

Berdasarkan Tabel 5.5 pada klaster 1 menunjukkan negara dengan tingkat penyebaran penyakit TBC rendah dengan nilai rata-rata pada variabel TB/HIV sebesar 0,0361%, variabel TB Paru dengan rata-rata 0,0411%, variabel EPTB dengan rata-rata 0,0308%, variabel anak-anak sebesar 0,0042%, variabel perempuan sebesar 0,0172% dan pada variabel laki-laki sebesar 0,0274%. Kemudian pada klaster 2 menunjukkan negara dengan tingkat penyebaran

penyakit TBC tinggi dengan nilai rata-rata variabel TB/HIV sebesar 0,213%, variabel TB Paru dengan rata-rata 0,3887%, variabel EPTB dengan rata-rata 0,1789%, variabel anak-anak sebesar 0,0217%, variabel perempuan sebesar 0,1414% dan pada variabel laki-laki sebesar 0,259%. Kemudian pada klaster 3 menunjukkan negara dengan tingkat penyebaran penyakit TBC sedang dengan nilai rata-rata variabel TB/HIV sebesar 0,0536%, variabel TB Paru dengan rata-rata 0,2027%, variabel EPTB dengan rata-rata 0,0953%, variabel anak-anak sebesar 0,0284%, variabel perempuan sebesar 0,0816% dan pada variabel laki-laki sebesar 0,1174%. Terakhir pada Noise hanya terdapat satu negara yaitu Negara India dengan nilai penyebaran pada variabel TB/HIV sebesar 0,8917%, TB Paru dengan nilai 1,0156%, EPTB 0,706%, anak-anak sebesar 0,0743%, perempuan 0,4211% dan pada variabel laki-laki 7431. Hal ini menunjukkan bahwa hanya terdapat satu negara yang tergolong dalam noise mempunyai tingkat penyebaran penyakit TBC yang jauh lebih tinggi jika dibandingkan dengan klaster 1, klaster 2 maupun klaster 3.



BAB VI

PENUTUP

4.1 Kesimpulan

Berdasarkan hasil tahapan analisis yang telah dilakukan, diperoleh beberapa kesimpulan antara lain:

1. Gambaran pada data HBC menunjukkan bahwa rata-rata dari variabel TB/HIV memiliki nilai sebesar 0,0805% dari populasi penduduk negara-negara yang terdaftar dalam HBC, variabel TB Paru sebesar 0,1123% dari populasi, variabel EPTB sebesar 0,0701%, variabel penderita anak-anak sebesar 0,0096%, variabel perempuan sebesar 0,0451% dan variabel laki-laki sebesar 0,0759%. Terdapat 3 negara dengan jumlah penderita penyakit tertinggi berdasarkan enam variabel yang digunakan yaitu Negara India, China dan Indonesia.
2. Berdasarkan hasil kelompok dengan menggunakan metode DBSCAN, didapatkan hasil cluster sebanyak 3 kelompok, sedangkan terdapat 1 negara dianggap sebagai *noise*. Berdasarkan hasil pengelompokan pada klaster 1 terdapat 23 negara yang merupakan negara dengan penyebaran penyakit TBC rendah. Klaster 2 terdapat 2 negara yang merupakan negara dengan penyebaran penyakit TBC tinggi. Klaster 3 terdapat 2 negara yang merupakan negara dengan penyebaran penyakit TBC sedang.

4.2 Saran

Penelitian selanjutnya dapat membandingkan metode DBSCAN dengan metode *clustering* yang lainnya untuk mendapatkan metode *clustering* terbaik.

DAFTAR PUSTAKA

- Bäcklund, H., A. Hedblom, and N. Neijman. (2011). "A Density-Based Spatial Clustering of Application with Noise," Linköpings Universitet.
- Bharati, M. R. (2010). *DATA MINING TECHNIQUES AND APPLICATIONS*, 1-2.
- Bramer, M. (2013). *Principle of Data Mining Second Edition*. London: Springer.
- Chanafi, M., Hapsari, D. P., Hapsari, R. K., dan Indriyani, T. (2019). Implementasi Algoritma Clustering Untuk Pengelompokan Pelanggan Retail Berdasarkan Skor Recency, Frequency, Dan Monetary. 1-62.
- Devi, N. M., Putra, I. K., dan Sukarsa, I. M. (2015). Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan. *Lontar Komputer*, 6.
- Fu, Y. (1997). *What is data mining*. *Data Mining*, 18-20.
- Han, J., Hedblom, A., and Neijman, N. (2011). A Density-Based Spatial Clustering of Application with Noise.
- Han, J., Kamber, M., and Pei, J. (2001). *Data Mining Concepts and Technique*.
- Heo, J.-N., Choi, Y. W., Jeon, S. C., and Park, C. K. (2005). Pulmonary Tuberculosis: Another Disease Showing Klusters of Small Nodules. *American Journal of Roenthenology*, 184.
- Hair, J.F.Jr., R.E. Anderson, R.L. Thatham dan Black, W.C. 2010. *Multivariat Data Analysis Seventh Edition*. New Jersey: Pearson Education, Inc.
- Indah, M. (2018). *Infodatin*. (D. Bujianto, Producer) Retrieved from Pusdatin.kemkes.go.id:
<https://pusdatin.kemkes.go.id/resources/download/pusdatin/infodatin/infodatin-tuberkulosis-2018.pdf>
- Intaran, N. P. (2012). *Tugas Data Mining*. Program Studi Teknik Informatika Jurusan Ilmu Komputer FMIPA Universitas Udayana.

- Kristanto, A., Sedyono, E., dan Hartomo, K. D. (2020). Implementation DBSCAN Algorithm to Clustering Satellite Surface Temperature Data in Indonesia.
- Larose, D. T., dan Larose, C. D. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining, Second Edition*. USA: John Wiley dan Sons Inc.
- Manning, C. D., Raghavan, P., dan Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Misnadiarly. (2006). *Diabetes Melitus Gangren, Ulcer, Infeksi, Mengenal gejala, Menanggulangi, dan Mencegah komplikasi*. Jakarta: Pustaka Obor Populer. Jakarta: Pustaka Obor Populer.
- Putra, O. A. (2012). Mycobacterium Tuberculosis yang Resisten Terhadap Antibiotik Lini Pertama pada Pasien Tuberkulosis di RSUP Fatmawati. *Skripsi*.
- Rohman, H., dan Utami, N. R. (2019). Pengelompokan Kaus Tuberkulosis Paru dan Upaya Promosi Kesehatan Di Wilayah Mlati II Sleman Yogyakarta. *Jurnal Ilmu Kesehatan Masyarakat (Journal of Public Health Sciences)*, 8.
- Saputera, M. J. (2019). Implementasi Clustering pada Algoritma DBSCAN untuk Deteksi Kejadian Bencana Alam Pada Microblogging Twitter. 71-76.
- Sarwani, D., dan Sri Nurlaela, I. Z. (2012). Faktor Risiko Multidrug Resistant Tuberculosis (MDR-TB). *Jurnal Kesehatan Masyarakat*.
- Serin, A. (2011). *Biclustering Analysis for Large Scale Data*. Jerman: Universitas Berlin.
- Silitonga, P. (2017). Analisis Pola Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Dengan Menggunakan Metode Dbscan Clustering. *Jurnal TIMES*, Vol. V No 1 : 36-39, 2016.
- Supranto, J. (2004). *Analisis Multivariat Arti dan Interpretasi*. Jakarta: Rineka Cipta.
- Wardani, D. W., Lazuardi, L., Mahendradhata, Y., & Kusnanto, H. (2010). Pentingnya Analisis Klaster Berbasis Spasial dalam Penanggulangan Tuberkulosis di Indonesia.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Kaufmann Publishers.



LAMPIRAN



Lampiran 1 Script Program Software R

```
anas<-read.delim(pipe("pbpaste"), sep="\t", header= TRUE)
View(anas)
summary(anas)
attach(anas)
cor(anas[,2:5])

#fungsi uji multikolinieritas
#uji bartlett
uji_bart <- function(x)
{
  method <- "Bartlett's test of sphericity"
  data.name <- deparse(substitute(x))
  x <- subset(x, complete.cases(x))
  n <- nrow(x)
  p <- ncol(x)
  chisq <- (1-n+(2*p+5)/6)*log(det(cor(x)))
  df <- p*(p-1)/2
  p.value <- pchisq(chisq, df, lower.tail=FALSE)
  names(chisq) <- "Khi-squared"
  names(df) <- "df"
  return(structure(list(statistic=chisq, parameter=df,
p.value=p.value, method=method, data.name=data.name),
class="htest"))
}

uji_bart(anas[,2:5]) #H0 :p11 =p12 =...=p78 = 0

#Uji KMO
library(psych)
kmo <- function(x)
{
  x <- subset(x, complete.cases(x))
  r <- cor(x)
  r2 <- r^2
  i <- solve(r)
  d <- diag(i)
  p2 <- (-i/sqrt(outer(d, d)))^2
  diag(r2) <- diag(p2) <- 0
  KMO <- sum(r2)/(sum(r2)+sum(p2))
  MSA <- colSums(r2)/(colSums(r2)+colSums(p2))
  return(list(KMO=KMO, MSA=MSA))
}

kmo(anas[,2:5]) #sampel cukup untuk digunakan jika kmo>0.5

##Analisis PCA##
library(factoextra)
pca_anas<- prcomp(anas[,2:5], scale. = TRUE)
pca_anas
round(pca_anas$rotation,2)
#bayaknya jumlah faktor dilihat dari nilai eigen value > 1 atau
nilai varians > 80%
round(pca_anas$sdev^2,2)
fviz_eig(pca_anas, addlabels = TRUE, ylim = c(0, 80))
```



```

summary(pca_anas)
##data hasil PCA
pca_anas$x[,1:5]

library("fpc")
library("dbscan")

wahyu1<- data.frame(anas[,2:5])
dim(wahyu1)
kNNdistplot(wahyu1, k=2)
abline(h=13, lty=3)
dbscan1 <- fpc::dbscan(wahyu1, eps = 13, MinPts =2)
print(dbscan)
# Plot DBSCAN results
library("factoextra")
fviz_cluster(dbscan1, data = wahyu, stand = FALSE,outlier.color =
"red",ellipse = FALSE, shape = NULL, labelsize = 12, geom
=c("point", "text"),
           palette = "jco", ggtheme = theme_bw())
#Cluster membership. Noise/outlier observations are coded as 0
dbscan1$cluster

```



Lampiran 2 Data Tuberkulosis menurut HBC tahun 2018 (dalam persentase populasi di setiap negara)

Negara	TB/HIV	TB Paru	EPTB	Anak anak	Perempuan	Lakilaki
Bangladesh	0.0017	0.1344	0.1195	0.0066	0.0680	0.0913
Brazil	0.0404	0.0445	0.0379	0.0015	0.0148	0.0348
China	0.2964	0.4692	0.1828	0.0049	0.1531	0.3359
Congo	0.0633	0.0875	0.0812	0.0116	0.0411	0.0527
Ethiopia	0.0649	0.0487	0.0438	0.0071	0.0282	0.0353
India	0.8917	1.0156	0.7060	0.0743	0.4211	0.7431
Indonesia	0.1296	0.3082	0.1751	0.0385	0.1296	0.1821
Kenya	0.0575	0.0499	0.0341	0.0059	0.0188	0.0341
Mozambique	0.0562	0.0534	0.0224	0.0075	0.0241	0.0258
Myanmar	0.0763	0.0780	0.0377	0.0163	0.0249	0.0446
Nigeria	0.0645	0.0620	0.0497	0.0052	0.0219	0.0374
Pakistan	0.0448	0.1791	0.1075	0.0291	0.0940	0.1008
Philippines	0.0623	0.2262	0.0831	0.0277	0.0693	0.1339
Russia	0.0462	0.0447	0.0262	0.0015	0.0141	0.0331
South Africa	0.1275	0.1260	0.0991	0.0099	0.0524	0.0793
Thailand	0.0417	0.0449	0.0312	0.0005	0.0164	0.0359
Tanzania	0.0459	0.0367	0.0223	0.0065	0.0153	0.0246
Vietnam	0.0526	0.0495	0.0433	0.0012	0.0167	0.0439
Camboja	0.0167	0.0116	0.0094	0.0034	0.0064	0.0080
Central African Republic	0.0053	0.0053	0.0043	0.0011	0.0024	0.0032
Congo	0.0013	0.0051	0.0033	0.0005	0.0027	0.0035

Lesotho	0.0042	0.0039	0.0029	0.0002	0.0015	0.0027
Liberia	0.0037	0.0032	0.0029	0.0007	0.0018	0.0023
Namibia	0.0048	0.0039	0.0041	0.0004	0.0016	0.0028
Papua New Guinea	0.0090	0.0097	0.0052	0.0042	0.0062	0.0069
Sierra Leone	0.0104	0.0098	0.0069	0.0015	0.0035	0.0056
Zambia	0.0207	0.0190	0.0122	0.0013	0.0070	0.0135
Zimbabwe	0.0147	0.0139	0.0085	0.0009	0.0056	0.0091



