

# PENGELOMPOKAN DAN PEMETAAN PENYAKIT TUBERKULOSIS PARU MENURUT PROVINSI DI INDONESIA TAHUN 2016 MENGGUNAKAN ANALISIS CLUSTER K-MEANS

Arif Anjang Laksono<sup>1</sup>, Bana Ali Fikri<sup>1</sup>, Muhammad Atma Yadin<sup>1</sup>, Sendhyka Cakra  
Pradana<sup>1</sup>, Tegar Anugrah Widi<sup>1</sup>, Edy Widodo<sup>1</sup>

<sup>[1]</sup>Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas  
Islam Indonesia

<sup>[1]</sup>[14611071@students.uii.ac.id](mailto:14611071@students.uii.ac.id), [edywidodo@uui.ac.id](mailto:edywidodo@uui.ac.id)

## Abstrak

*Tuberkulosis (TBC atau TB) adalah suatu penyakit infeksi yang disebabkan oleh bakteri Mycobacterium tuberculosis. Penyakit ini paling banyak menyerang paru-paru. Indonesia setiap tahun ditemukan 582.000 kasus penderita baru TB dengan angka kematian 41 orang per 100.000 sebagian besar penderita TB paru atau sebanyak 75% paling mencengangkan yang terkena virus TB paru ialah orang-orang yang masi usia produktif 15-49 tahun. K-Means merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering suatu data. Ada banyak pendekatan untuk membuat cluster, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam kelompok yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Tujuan penelitian ini untuk mengetahui cluster penyakit Tuberculosis Paru menurut provinsi di Indonesia dan pemetaanya berdasarkan kelompok umur penderita Tuberculosis. Untuk penyelesaian permasalahan yang ada maka penulis akan menggunakan analisis cluster K-Means untuk mengetahui kelompok daerah yang memiliki kasus TB yang sama, sehingga dapat diketahui wilayah yang akan menjadi prioritas dalam menangani kasus TB. Penerapan kasus Tuberkulosis yang dianalisis menggunakan cluster k-means menjadi tiga cluster. Cluster 1 merupakan provinsi dengan jumlah penderita rendah yang terdiri dari 29 Provinsi yaitu Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Banten, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Sulawesi Selatan, Maluku, Maluku Utara, Papua Barat dan Papua. Cluster 2 merupakan provinsi dengan jumlah penderita sedang yang terdiri dari 4 Provinsi yaitu Sumatera Utara, DKI Jakarta, Jawa Tengah dan Jawa Timur dan cluster 3 merupakan provinsi dengan jumlah penderita tinggi yang terdiri 1 Provinsi yaitu Jawa Barat dan hasil pemetaan provinsi berdasarkan penderita tuberculosis dari kelompok umur.*

**Kata Kunci :** Tuberkulosis; Clustering; K-means; Provinsi.

## 1. PENDAHULUAN

Tuberkulosis adalah penyakit radang parenkim paru karena infeksi kuman *Mycobacterium tuberculosis*. Tuberkulosis paru termasuk suatu pneumonia, yaitu pneumonia yang disebabkan oleh *M. tuberculosis* (Darmanto, 2014) dikutip (Yoanisa Aprilia, 2017).

Dalam laporan WHO tahun 2013 diperkirakan terdapat 8.6 juta kasus Tuberkulosis pada tahun 2012 dimana 1.1 juta orang (13%) diantaranya adalah

padien dengan HIV positif. Sekitar 75% dari pasien tersebut berada di wilayah Afrika. Pada tahun 2012 diperkirakan terdapat 450.000 orang yang menderita Tuberkulosis MDR dan 170.000 diantaranya meninggal dunia.

Di Indonesia *Tuberculosis* (TB) merupakan salah satu penyakit yang menimbulkan masalah kesehatan di masyarakat penderita TB di Indonesia peringkat ke-3 setelah India dan China dengan total penderita 10% di dunia diperkirakan tahun 2004 ada 539.000 kasus baru dan kematian 101.000 orang. Insiden kasus TB sekitar 110 per 100.000 penduduk (Depkes, 2007).

Sebelum dilakukan penelitian tentang pengelompokan penyakit tuberkulosis, penulis merujuk pada penelitian sebelumnya oleh Noor Anisha Suherni dan Maduratna (2013) tentang analisis pengelompokan kecamatan di Kota Surabaya berdasarkan faktor penyebab terjadinya penyakit TB dan penelitian dari Windha Mega Pradnya Duhita (2015) tentang *clustering* menggunakan metode *k-means* untuk menentukan status gizi balita.

*K-Means* merupakan salah satu algoritma dalam *data mining* yang bisa digunakan untuk melakukan pengelompokan/*clustering* suatu data. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam kelompok yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah *clustering* (Witten et al 2013).

Untuk penyelesaian permasalahan yang ada maka penulis akan menggunakan analisis *cluster K-Means* untuk mengetahui kelompok daerah yang memiliki kasus TB yang sama, sehingga dapat diketahui wilayah yang akan menjadi prioritas dalam menangani kasus TB.

Dari latar belakang yang sudah dipaparkan maka tujuan dari penelitian ini adalah untuk melihat gambaran umum dan pengelompokan penyakit TB menurut provinsi di Indonesia tahun 2016.

## 2. METODE PENELITIAN

### 2.1 Populasi

Populasi yang digunakan dalam penelitian adalah semua penduduk yang ada di 34 provinsi di Indonesia yang menderita penyakit TB pada tahun 2016

### 2.2 Variabel dan Definisi Operasional Variabel

Tabel 1. Variabel dan Definisi Operasional Variabel

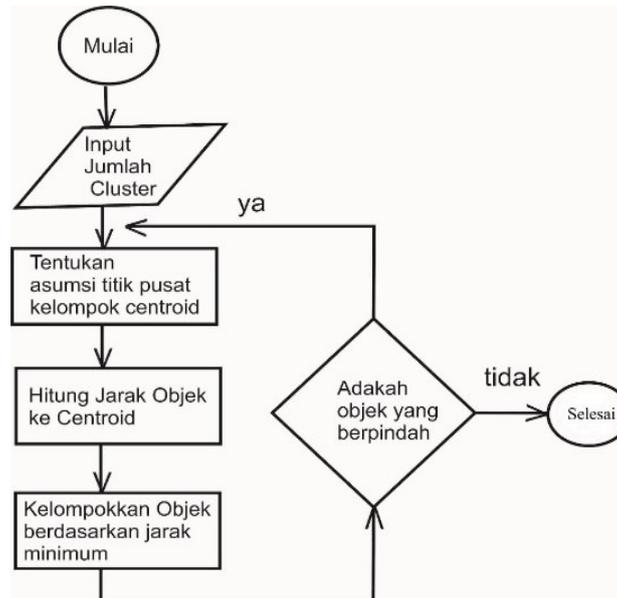
Variabel	Definisi Operasional Variabel
Provinsi	34 provinsi yang ada di Indonesia
Umur	Rentang umur penderita penyakit tuberkulosis (0-14), (15-24), (25-34), (35-44), (45-54), (55-64), ( $\geq 65$ )
Total Penderita Tuberkulosis	Jumlah keseluruhan penderita penyakit tuberkulosis

### 2.3 Metode Pengambilan Data

Data yang digunakan merupakan data sekunder yang bersumber dari Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia, yang ada di website Departemen Kesehatan Republik Indonesia.

## 2.4 Tahapan Penelitian

Tahapan dari penelitian ini dimulai dengan menentukan besarnya  $k$ , yaitu banyaknya *cluster* yang diinginkan dan menentukan *centroid* di tiap *cluster*.



Gambar 1. Flowchart Alur Penelitian

Kedua menghitung jarak tiap objek dengan pusat *centroid* yang telah ditentukan.

Ketiga menghitung kembali rata-rata (*centroid*) untuk *cluster* yang baru terbentuk. Setelah itu mengulangi langkah 2 sampai tidak ada lagi pemindahan objek antar *cluster*.

Setelah *cluster* terbentuk, baik dengan metode hirarki maupun non hirarki, langkah selanjutnya melakukan interpretasi terhadap *cluster* yang terbentuk, yang pada intinya memberi nama spesifik untuk menggambarkan isi *cluster* tersebut

## 3. HASIL PENELITIAN DAN PEMBAHASAN

### Analisis Deskriptif

Dari data yang digunakan dalam penelitian ini, diperoleh gambaran (deskriptif) dan grafik pada masing-masing variabel umur jumlah penderita penyakit tuberkulosis di Indonesia pada tahun 2016.

Tabel 2. Deskriptif Data Penderita Penyakit Tuberkulosis

Kelompok Umur (Tahun)	N	Minimum	Maksimum	Mean	Std. Deviation
X <sub>1</sub> (0-14)	34	2	200	44.32	49.773
X <sub>2</sub> (15-24)	34	79	4972	752.59	990.806
X <sub>3</sub> (25-34)	34	104	5392	890.79	1102.368
X <sub>4</sub> (35-44)	34	116	5066	907.47	1062.143
X <sub>5</sub> (45-54)	34	119	4819	911.59	1061.695
X <sub>6</sub> (55-64)	34	90	4506	881.32	1038.614
X <sub>7</sub> (>=65)	34	84	4602	915.41	1115.283

Pada tabel 2 dapat dilihat jumlah penderita penyakit tuberkulosis tertinggi yaitu penderita dengan kelompok umur 25 - 34 tahun dengan total jumlah penderita sebanyak 5392 penderita sedangkan kelompok umur 0 - 14 tahun merupakan kelompok dengan jumlah penderita paling sedikit yaitu berjumlah 2 penderita. Pada analisis deskriptif ini juga didapatkan hasil rata-rata penderita tuberkulosis dari masing-masing variabel dan variabel yang mempunyai rata-rata tertinggi yaitu variabel x7 atau kelompok umur 65 tahun keatas dengan rata-rata sebesar 915,41 sedangkan yang paling rendah adalah kelompok umur 0-14 tahun atau variabel x1 dengan rata-rata sebesar 44,32.

Selain itu pada analisis deskriptif ini juga didapatkan nilai std. Deviasi yang nantinya digunakan untuk mencari nilai rata-rata variabel antar *cluster*, akan tetapi terlebih dahulu data yang digunakan ditransformasikan terlebih dahulu ke bentuk nilai yang lebih kecil untuk memudahkan dalam mencari rata-rata variabel antar *cluster*.

### Standardisasi (Transformasi)

Data yang terkumpul mempunyai variabilitas satuan yang berbeda, maka perlu dilakukan langkah standardisasi atau transformasi terhadap variabel yang relevan ke bentuk *z-score*.

**Tabel 3.** Hasil Z-Score Penderita Tuberkulosis

Provinsi	zx1	zx2	zx3	zx4	zx5	zx6	zx7
ACEH	-0,60924	-0,35687	-0,30915	-0,32902	-0,25298	-0,20732	-0,20928
SUMATRA UTARA	0,85743	0,98244	1,18400	1,26304	1,34823	1,36401	1,43066
SUMATRA BARAT	1,19898	-0,16511	-0,17217	-0,17179	-0,19553	-0,22176	-0,22094
RIAU	0,09396	-0,21052	-0,15131	-0,13319	-0,07120	-0,07734	-0,10169
JAMBI	-0,38824	-0,46890	-0,45610	-0,44012	-0,43006	-0,42492	-0,40475
SUMATRA SELATAN	0,25469	-0,03087	0,10088	0,20198	0,24057	0,21055	0,18792
BENGKULU	-0,64942	-0,62635	-0,65749	-0,68114	-0,68531	-0,68006	-0,64953
LAMPUNG	-0,02659	-0,19034	-0,10776	-0,04375	-0,03729	-0,04268	-0,02906

Pada tabel 3 dapat dilihat nilai hasil transformasi dari data awal ke bentuk data yang sudah ditransformasi pada masing-masing variabel sehingga didapatkan hasil titik pusat *cluster* terakhir

**Tabel 4.** Final Cluster Centers

	Final Cluster Centers		
	1	2	3
Zscore(x1)	-0,36814	1,03825	3,07752
Zscore(x2)	-0,37597	1,02282	3,218
Zscore(x3)	-0,38574	1,12073	3,15884
Zscore(x4)	-0,38972	1,17501	3,10601
Zscore(x5)	-0,39077	1,15962	3,15148
Zscore(x6)	-0,39184	1,15243	3,18085
Zscore(x7)	-0,38784	1,11841	3,19299

Hasil pada tabel 4 didapatkan dari hasil perhitungan menjumlahkan objek ke *i* pada variabel *j* yang dikurangkan dengan pusat kelompok ke-*k* pada variabel *j*

$$d_{ik} = \sqrt{\sum_{j=1}^m (X_{ij} - C_{kj})^2} \quad \dots (1)$$

Maka didapatkanlah hasil seperti pada tabel ...yang nantinya akan digunakan untuk mencari rata-rata variabel antar *cluster*

### Menghitung Rata-Rata Variabel Antar *cluster*

Rumus umum yang digunakan yaitu:

$$X = \mu + z \cdot \sigma \dots (2)$$

dengan:

X : rata-rata sampel (variabel dalam *cluster*)

$\mu$  : rata-rata populasi

z : nilai standarisasi

$\sigma$  : standar deviasi

Untuk mengetahui rata-rata jumlah variabel x1 pada *cluster*-1 yaitu :

$$X = (\text{rata-rata jumlah pada variabel x1}) + (-0,36814 \times \text{standar deviasi variabel x1})$$

$$= 44,32 + (-0,36814 \times 49,773)$$

$$= 26,00$$

Dari hasil perhitungan didapatkan hasil titik pusat Cluster terakhir. Hasil perhitungan pusat Cluster terakhir ini akan digunakan untuk *profilisasi* pada tiap – tiap *cluster* yang terbentuk.

### Final Cluster

*Output Final Cluster* pada tabel 5 merupakan hasil perhitungan *final* jarak titik pusat cluster

Tabel 5. *Final Cluster Centers*

Final Cluster Centers			
	Cluster		
	1	2	3
x1	26,000	96	197,5
x2	380,075	1766	3941
x3	465,567	2126,25	4373
x4	493,532	2155,5	4206,5
x5	496,710	2142,75	4257,5
x6	474,353	2078,25	4185
x7	482,860	2162,76	4476,5

*Output Final Cluster* akan digunakan untuk *profilisasi* pada tiap – tiap *cluster* yang terbentuk dengan cara membandingkan rata-rata variabel di tiap-tiap *cluster*.

## Iterasi

Tabel 6. Iteration History

Iteration	Change in Cluster Centers		
	1	2	3
1	0.964	0.499	0.000
2	0.083	0.606	0.000
3	0.000	0.000	0.000

\*jarak minimum antar pusat kluster 5,610.

Dari hasil output iterasi yang dicari dengan bantuan software statistika menunjukkan proses *clustering* dilakukan melalui tiga tahapan iterasi untuk mendapatkan cluster yang lebih akurat. Dapat diketahui bahwa jarak minimum antar pusat *cluster* yang terjadi dari hasil iterasi adalah sebesar 5,610.

## Anggota Cluster

Dari hasil *cluster* yang telah dilakukan didapatkan anggota pada masing masing *cluster*. *Cluster 1* berjumlah 29 anggota atau 29 Provinsi yaitu Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Banten, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Sulawesi Selatan, Maluku, Maluku Utara, Papua Barat dan Papua.

Pada Cluster 2 terdapat 4 anggota Provinsi yaitu Sumatera Utara, DKI Jakarta, Jawa Tengah dan Jawa Timur.

Pada *Cluster 3* didapatkan *Cluster* dengan provinsi Jawa Barat sebagai anggota Cluster 3

## Pemetaan



Gambar 2. Hasil Pengelompokan Cluster

Berdasarkan hasil pengelompokan yang mengelompokkan provinsi menjadi 3 kelompok (*cluster*) gambar peta diatas adalah hasil dari pemetaan pada setiap

kelompok. Pembagian warna dari warna muda ke tua digunakan untuk mempermudah visualisasi peta. Warna coklat muda digunakan untuk kelompok *cluster* 1, warna coklat digunakan untuk kelompok *cluster* 2 dan terakhir warna coklat tua digunakan untuk kelompok *cluster* 3.

#### 4. SIMPULAN

Dari hasil pembahasan sebelumnya, dapat diperoleh kesimpulan, yaitu :

4.1 Hasil cluster terbentuk menjadi tiga cluster.

4.2 Berikut anggota dan ciri – ciri dari masing-masing cluster:

- a. *Cluster* 1 merupakan provinsi dengan jumlah penderita rendah karena selain jumlah penduduk yang tidak terlalu besar serta pabrik rokok dan industri-industri yang tidak banyak sehingga pencemaran udaranya juga kecil. Cluster 1 terdiri dari 29 Provinsi yaitu Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Banten, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalbar, Kalteng, Kalsel, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Sulawesi Selatan, Maluku, Maluku Utara, Papua Barat dan Papua.
- b. *Cluster* 2 merupakan Provinsi dengan jumlah penderita menengah atau sedang karena pabrik serta industri dan pencemaran udaranya tidak terlalu besar. Cluster 2 terdiri dari 4 Provinsi yaitu Sumatera Utara, DKI Jakarta, Jawa Tengah dan Jawa Timur.
- c. *Cluster* 3 merupakan provinsi dengan jumlah penderita penyakit TB terbanyak karena jumlah penduduk di Provinsi Jawa Barat tinggi dan banyak pabrik rokok di pulau Jawa banyak industri sehingga banyak pencemaran udara cluster 3 terdiri 1 Provinsi yaitu Jawa Barat.

#### 5. DAFTAR PUSTAKA

- Admin. 2017. Tuberculosis. [www.who.int/mediacentre/factsheets/fs104/en/](http://www.who.int/mediacentre/factsheets/fs104/en/) (Diakses 10 Oktober 2017 Pukul 18.45)
- Binus. 2017. Clustering. <https://socs.binus.ac.id/2017/03/09/clustering/> (Diakses pada 25 September 2017 Pukul 09.00)
- Depkes R.I., 2008. **Profil Kesehatan Indonesia**. Jakarta
- Depkes R.I (2009). Profil Kesehatan Indonesia. Jakarta : Departemen Republik Indonesia.
- Kemenkes R.I. 2016. “Data dan Informasi Profil Kesehatan Indonesia”. Dalam <http://www.depkes.go.id>. Download 24 September 2017 Pukul 14.00 WIB.
- Kuriah, Dwi. (2016). Laporan Kerja Praktek “*Analisis Cluster K-Means untuk Mengelompokkan Budidaya Produksi Ikan dan Segmentasi Produksi Ikan di Kabupaten Sragen*”. Laporan Kerja Praktek, Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia Yogyakarta.
- Metisen, Benri Melpa dan Herlina Latipa Sari. (2015). Jurnal “*Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila*”. Program Studi Teknik Informatika Fakultas Ilmu Komputer Univesitas Dehasen Bengkulu. Bengkulu.
- Mutianingtyas, Yoanisa Aprilia. 2017. “Hubungan Perilaku Pasien Tuberkulosis Paru dengan Kepatuhan Program Pengobatan Sistem DOTS di Balai

- Kesehatan Paru Masyarakat Purwokerto”. Skripsi. FIK, Ilmu Keperawatan, Universitas Muhammadiyah Purwokerto.
- Osgeo. 2013. “Mengenal SIG dan data Spasial”. [http://osgeo.ft.ugm.ac.id/mengenal-sig-dan-data-spasi al/](http://osgeo.ft.ugm.ac.id/mengenal-sig-dan-data-spasi-al/) (Diakses pada 25 September 2017 Pukul 09.40).
- Pradnya, Mega Wendha. 2015. “*Clustering Menggunakan Metode K-means untuk Menentukan Status Gizi Balita*”. Jurnal
- Setiawan, Yohanes. (2012). Makalah “*Rancang Bangun Aplikasi Pemetaan Layanan Rumah Sakit Darmo Surabaya dengan Menggunakan Teknologi Layar Sentuh.*” Makalah, Jurusan Sistem Informasi Fakultas Teknologi dan Informatika Sekolah Tinggi Manajemen Informatika dan Teknik Komputer Surabaya. Surabaya.
- Statisticsanalyst. 2011. “Metode Spasial”. <https://statisticsanalyst.wordpress.com/2011/08/26/metode-spasial/> (Diakses pada 25 September 2017 Pukul 09.30).
- Suherni, Noor Anisa Dewi dan Maduratna. (2013). Jurnal ”*Analisis Pengelompokan Kecamatan di Kota Surabaya Berdasarkan Faktor Penyebab Terjadinya Penyakit Tuberkulosis*”. Jurusan Statistika Fakultas MIPA ITS. Surabaya.
- Yoga F. 2015. Faktor- Faktor yang Mempengaruhi Kejadian TB Paru dan Upaya Penang gulangannya.