

PENGELOMPOKAN JUMLAH KASUS TUBERCULOSIS PARU DI INDONESIA MENGGUNAKAN CLUSTER K-MEANS TAHUN 2016

Andi Nurhanna Manthovani¹, Septi Serdawati², Cindy Fatika Sari³, Ika Fitia Widiawati⁴, Zarmeila Putri⁵, Edy Widodo⁶

^{1,2,3,4,5}Mahasiswa Program Studi Statistika Universitas Islam Indonesia

⁶Dosen Program Studi Statistika Universitas Islam Indonesia

¹14611182@students.uii.ac.id, ²14611183@students.uii.ac.id,

³14611192@students.uii.ac.id, ⁴14611193@students.uii.ac.id,

⁵14611194@students.uii.ac.id, ⁶edywidodo.uii.ac.id

Abstrak

Tujuan dari penelitian ini yaitu untuk mengetahui gambaran umum TB Paru, mengetahui pengelompokan TB Paru dan karakteristik sebaran TB Paru menurut Persentase Jumlah Kasus, Persentase Penderita Sembuh, Persentase Jumlah Penduduk dan Provinsi di Indonesia tahun 2016. Alat yang digunakan dalam penelitian ini menggunakan cluster K-Means. Berdasarkan penelitian menggunakan software R diperoleh 4 kelompok yang mempunyai karakteristik mirip berdasarkan provinsi di Indonesia. Kelompok 1 mempunyai karakteristik provinsi-provinsi dengan persentase kesembuhan penderita TB paru terbanyak kedua, persentase jumlah kasus yang paling sedikit dan persentase yang paling banyak. Kemudian kelompok 2 mempunyai karakteristik provinsi-provinsi dengan persentase kesembuhan penderita TB paru terkecil kedua, persentase kasus yang lumayan banyak dan persentase jumlah penduduk yang lumayan sedikit. Selanjutnya kelompok 3 dengan karakteristik provinsi-provinsi dengan persentase kesembuhan penderita TB paru terbesar, persentase jumlah kasus yang kecil dan persentase jumlah penduduk yang banyak. Kemudian kelompok terakhir provinsi-provinsi dengan persentase kesembuhan penderita TB paru terkecil, persentase jumlah kasus yang banyak dan persentase jumlah penduduk yang kecil yaitu dibawah 2%.

Kata Kunci: Bakteri; kelompok ;K-Means;TB

1. PENDAHULUAN

1.1 Latar Belakang

TB adalah suatu penyakit infeksi yang disebabkan oleh bakteri *mycobacterium tuberculosis*. Bakteri ini merupakan bakteri basil yang sangat kuat sehingga memerlukan waktu lama untuk mengobatinya. Bakteri ini lebih sering menginfeksi organ paru-paru dibandingkan bagian lain tubuh manusia.

Hasil survei Kesehatan Rumah Tangga Depkes RI tahun 1992, menunjukkan bahwa TB merupakan penyakit kedua penyebab kematian, sedangkan pada tahun 1986 merupakan penyebab kematian keempat. Pada tahun 1999 WHO memperkirakan di Indonesia terdapat 583.000 penderita TB paru pertahun dengan 262.000 BTA positif atau insidens rate kira-kira 130 per 100.000 penduduk. Kematian akibat TB diperkirakan menimpa 140.000 penduduk tiap tahun. Berdasarkan data WHO Global Tuberculosis Report 2016 menyatakan Indonesia menempati ranking kedua dengan jumlah kasus TB terbanyak di dunia.

Penderita TB dilaporkan meningkat secara drastis pada dekade terakhir ini di seluruh dunia. Demikian pula di Indonesia, dengan penduduk lebih dari 200 juta orang, Indonesia menempati urutan kedua pada tahun 2015 setelah India dengan penderita sebanyak lebih dari 1 juta kasus.

Dari sekian banyak kasus TB di Indonesia jumlah kasus TB paru yang paling banyak, karena TB paru cepat menular dibandingkan TB lainnya. Oleh karena itu, Penelitian kali tertarik meneliti TB paru di Indonesia pada tahun 2016 dengan variabel persentase jumlah kasus TB paru, persentase kasus sembuh dan persentase jumlah penduduk yang akan dikelompokkan berdasarkan provinsi di Indonesia. Selanjutnya untuk meneliti kasus penderita TB paru di Indonesia menggunakan *cluster K-means*. Sebagaimana diketahui *cluster K-means* merupakan suatu alat pengelompokan yang mudah untuk diadaptasi dan sangat umum digunakan.

Penelitian mengenai penyakit TB telah banyak dilakukan sebelumnya, salah satunya adalah penelitian pengelompokkan dan korelasi kasus TB Paru BTA positif dengan kepadatan penduduk dan kemiskinan penduduk yang ditinjau dengan menggunakan studi epidemiologi serta untuk melihat tren kasus TB Paru BTA positif berdasarkan penggunaan aplikasi web Sistem Informasi Geografis (Hastuti et al., 2015). Hasil dari penelitian tersebut menunjukkan tidak ada perbedaan yang signifikan antara kepadatan penduduk tinggi dan kepadatan penduduk rendah serta jumlah keluarga miskin tinggi dan jumlah keluarga miskin rendah terhadap kasus TB Paru BTA positif serta pemetaan hasil dari pengelompokkan penyakit TB Paru BTA Positif di Kota Kendari tahun 2013-2015.

Berdasarkan latar belakang diatas, maka peneliti tertarik melakukan penelitian yang berjudul “Pengelompokkan Jumlah Kasus *Tuberculosis* Paru di Indonesia Menggunakan *Cluster K-Means* Tahun 2016 ”

1.2 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka penelitian ini bertujuan untuk mengetahui pengelompokkan dan karakteristik sebaran TB Paru menurut Persentase Jumlah Kasus, Persentase Penderita Sembuh, Persentase Jumlah Penduduk berdasarkan Provinsi di Indonesia tahun 2016.

2. METODE PENELITIAN

2.1 Populasi dan Sampel Penelitian

Pada penelitian ini populasi sasaran adalah jumlah kasus baru penyakit TB Paru yang terjadi di Indonesia pada tahun 2016. Sementara itu, sampel yang diambil adalah berdasarkan menurut presentase jumlah penduduk, presentase jumlah kasus, dan presentase jumlah penderita sembuh yang terjadi di Indonesia pada tahun 2016.

2.2 Teknik Pengumpulan Data

Metode pengumpulan data dalam penelitian ini adalah Metode Dokumentasi (pencatatan data sekunder) yaitu metode pengumpulan data yang menggunakan dokumen atau catatan tertulis dari laporan Departemen Kesehatan Indonesia yaitu www.depkes.go.id.

2.3 Variabel dan Definisi Operasional Peubah

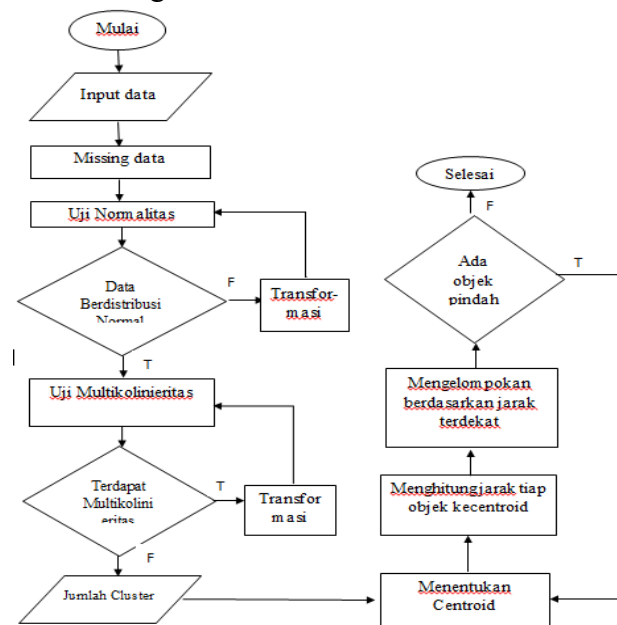
Tabel 2.7. Definisi Operasional Peubah

Variabel	Deskripsi	Skala Data
Presentase Jumlah Kasus	Menunjukkan presentase berapa banyak yang menderita penyakit TB Paru dari jumlah penduduk yang ada di provinsi di Indonesia	Nominal

Presentase Jumlah Penderita Sembuh	Menunjukkan berapa banyak yang sembuh dari penyakit TB Paru dari jumlah kasus TB Paru yang ada di provinsi di Indonesia	Nominal
Presentase Jumlah Penduduk	Menunjukkan presentase jumlah penduduk yang ada disetiap provinsi di Indonesia pada tahun 2016	Nominal
Provinsi	Menunjukkan provinsi dimana penderita penyakit kasus baru penyakit TB Paru berada.	-

2.4 Metode Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah penjabaran keadaan TB Paru secara umum dan pengelompokan data menggunakan metode *Clustering K-Means* dan pemetaan wilayah Indonesia sesuai dengan *Cluster* yang telah dibuat. Secara garis besar diagram alir dari penelitian ini adalah sebagai berikut.



Gambar 2.5 Diagram Alir Algoritma K-Means

2.5 Uji Normalitas

Menurut Santosa dan Ashari (2005, h231), pengujian normalitas adalah pengujian tentang kenormalan distribusi data. Uji ini merupakan pengujian yang paling banyak dilakukan untuk analisis statistik parametrik. Distribusi normal data dalam bentuk distribusi normal dimana data memusat pada nilai rata-rata dan median.

2.6 Transformasi Logaritma

Data yang tidak terdistribusi secara normal dapat ditransformasi agar berdistribusi normal. Namun sebelumnya, harus mengetahui terlebih dahulu bentuk grafik histogram dari data agar dapat menentukan bentuk transformasi data.

Transformasi Logaritma digunakan apabila data tidak memenuhi asumsi pengaruh aditif. Jika X adalah data asli, maka x' (X aksen) adalah data hasil transformasi, dimana $x' = \text{Log } x$.

2.7 Deteksi Multikolinearitas dengan Nilai VIF

Gujarati (1995) dalam Ifadah (2011). Istilah multikolinearitas mula-mula ditemukan oleh Ragnar Frisch. Multikolinearitas berarti adanya hubungan linear yang sempurna atau pasti, diantara beberapa atau semua variabel bebas dari model regresi ganda. Multikolinearitas biasanya terjadi ketika sebagian besar variabel yang digunakan saling terkait dalam suatu model regresi.

Beberapa ahli berpendapat bahwa nilai toleransi kurang dari 1 atau VIF lebih besar dari 10 menunjukkan multikolinearitas signifikan, sementara itu para ahli lainnya menegaskan bahwa besarnya R^2 model dianggap mengindikasikan adanya multikolinearitas. Klein (1962) menunjukkan bahwa, jika VIF lebih besar dari $1/(1 - R^2)$ atau nilai toleransi kurang dari $(1 - R^2)$, maka multikolinearitas dapat dianggap signifikan secara statistik.

2.8 Cluster K-Means

Algoritma *K-Means* diperkenalkan oleh J.B. MacQueen pada tahun 1976, salah satu algoritma *clustering* sangat umum yang mengelompokkan data sesuai dengan karakteristik atau ciri-ciri bersama yang serupa. Grup data ini dinamakan sebagai *cluster*. Data di dalam suatu *cluster* mempunyai ciri-ciri (atau fitur, karakteristik, atribut, properti) serupa dan tidak serupa dengan data pada cluster lain.

Algoritma *K-Means* merupakan salah satu algoritma dengan *partitional*, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* atau titik pusat *cluster* awalnya (Madhulatha 2012). Algoritma *K-Means* menggunakan proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan menghasilkan titik *centroid* akhir sebagai *output*. Metode *K-Means* akan memilih pola k sebagai titik awal *centroid* secara acak atau *random*. Jumlah iterasi untuk mencapai *clustercentroid* akan dipengaruhi oleh calon *clustercentroid* awal secara *random*. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan *centroidcluster* yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi (HUNG et al., 2005, Saranya & Punithavalli, 2011, Eltibi & Ashour, 2011).

Dalam penyelesaiannya, algoritma *K-Means* akan mengelompokkan item data dalam suatu dataset ke suatu *cluster* berdasarkan jarak terdekat (Bangoria et al., 2013). Nilai *centroid* awal yang dipilih secara acak yang menjadi titik pusat awal, akan dihitung jarak dengan semua data menggunakan rumus *Euclidean Distance*. Data yang memiliki jarak pendek terhadap *centroid* akan membuat sebuah *cluster*. Proses ini berkelanjutan sampai tidak terjadi perubahan pada setiap kelompok (Agrawal & Gupta, 2013, Chaturved & Rajavat, 2013). Algoritma *K-Means* memerlukan 3 komponen, yaitu :

1. Jumlah *Cluster K*

K-Means merupakan bagian dari metode *non-hirarki* sehingga dalam metode ini jumlah k harus ditentukan terlebih dahulu. Tidak terdapat aturan khusus dalam menentukan jumlah *cluster* k , terkadang jumlah *cluster* tergantung subyektif seseorang.

2. Cluster Awal

Cluster awal yang dipilih berkaitan dengan penentuan pusat *cluster* awal (*centroid* awal). Dalam hal ini terdapat beberapa pendapat dalam memilih *cluster* awal.

- Pemilihan *cluster* awal berdasarkan interval dari jumlah setiap observasi.
- Pemilihan *cluster* awal dapat ditentukan melalui pendekatan salah satu metode hirarki.
- Pemilihan *cluster* awal dapat secara acak dari semua observasi.

3. Ukuran Jarak

Metode *K-Means* dimulai dengan pembentukan *cluster* awal kemudian secara *iterative prototipe cluster* ini diperbaiki hingga konvergen (tidak terjadi perubahan yang signifikan) perubahan ini diukur dengan ukuran jarak. Ukuran jarak ini digunakan untuk menempatkan observasi ke dalam *centroid* terdekat.

- a. Kemudian dihitung jarak antara setiap objek dengan setiap pusat kelompok. Untuk melakukan penghitungan jarak objek ke- i pada pusat kelompok ke- k dapat digunakan rumus jarak *euclidean*, yaitu:

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

Keterangan:

- d_{ik} : jarak objek ke- i pada pusat kelompok ke- k
 x_{ij} : nilai objek ke- i pada variabel j
 c_{kj} : pusat kelompok ke- k pada variabel j
 m : jumlah variabel yang digunakan
 i menyatakan objek, k menyatakan kelompok
 j menyatakan keanggotaan kelompok

- b. Suatu objek akan menjadi anggota dari kelompok ke- k apabila jarak objek tersebut ke pusat kelompok ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat kelompok lainnya.
 c. Selanjutnya, kelompokkan objek-objek yang menjadi anggota pada setiap kelompok.
 d. Menentukan nilai pusat kelompok yang baru dapat dihitung dengan cara mencari nilai rata-rata dari objek yang menjadi anggota pada kelompok tersebut, dengan rumus sebagai berikut

$$c_{kj} = \frac{\sum_{h=1}^p y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke } - k \dots\dots\dots(2)$$

keterangan:

- c_{kj} : pusat kelompok ke- k pada variabel j
 y_{hj} : nilai objek ke- h pada variabel j
 p : jumlah anggota kelompok terbentuk

- e. Melakukan perulangan dari langkah 2 hingga langkah 5 hingga anggota tiap *cluster* tidak ada yang berubah.

- f. Jika langkah f telah terpenuhi, maka nilai pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

3. HASIL PENELITIAN DAN PEMBAHASAN

a. Uji Normalitas

Tabel 3.1 Uji Normalitas

One-Sample Kolmogorov-Smirnov Test			
	Persentase Jumlah Kasus	Persentase Jumlah Penderita Sembuh	Persentase Jumlah Penduduk
N	34	34	34
Kolmogorov-Smirnov Z	2.303	.789	1.691
Asymp. Sig. (2-tailed)	.000	.562	.007

a. Test distribution is Normal.

Berdasarkan **Tabel 3.1** dari hasil uji normalitas diperoleh nilai $P\text{-value}_{\text{kasus}} = 0.000$, $P\text{-value}_{\text{sembuh}} = 0.562$, $P\text{-value}_{\text{penduduk}} = 0.007$, sehingga dikarenakan $P\text{-value}_{\text{kasus}}$ dan $P\text{-value}_{\text{penduduk}} > \alpha$ maka $p\text{-value}$ lebih besar dari α sehingga gagal tolak H_0 . Sedangkan untuk $P\text{-value}_{\text{sembuh}} > \alpha$ maka $p\text{-value}$ lebih kecil dari α sehingga H_0 ditolak. Sehingga diperoleh kesimpulan dengan tingkat signifikansi 5% dan $p\text{ value} < \alpha$ maka H_0 ditolak yang artinya persentase jumlah kasus dan persentase jumlah penduduk tidak berdistribusi normal. Sedangkan dengan tingkat signifikansi 5% dan $p\text{ value} > \alpha$ maka gagal tolak H_0 yang artinya persentase jumlah penderita sembuh berdistribusi normal.

Berdasarkan uji normalitas yang dilakukan diketahui 2variabel tidak berdistribusi normal maka perlu mentransformasikan data. Maka uji normalitas kembali dilakukan seperti pada **Tabel 3.2**.

Tabel 3.3. Uji Normalitas

One-Sample Kolmogorov-Smirnov Test			
	Persentase Jumlah Penderita Sembuh	ln_jumlah_kasus	ln_jumlah_penduduk
N	34	34	34
Kolmogorov-Smirnov Z	.789	1.134	.631
Asymp. Sig. (2-tailed)	.562	.152	.820

a. Test distribution is Normal.

Berdasarkan **Tabel 3.2** dapat dilakukan uji normalitas pada data yang disediakan diperoleh $P\text{-value}_{\text{kasus}} = 0.152$, $P\text{-value}_{\text{sembuh}} = 0.562$, $P\text{-value}_{\text{penduduk}} = 0.820$. Dikarenakan $P\text{-value}_{\text{kasus}}$, $P\text{-value}_{\text{sembuh}}$, $P\text{-value}_{\text{penduduk}} < \alpha$ maka $p\text{-value}$ lebih kecil dari α sehingga H_0 ditolak. Sehingga diperoleh kesimpulan dengan tingkat signifikansi 5% dan $p\text{ value} < \alpha$ maka H_0 ditolak yang artinya persentase jumlah kasus, persentase jumlah penderita sembuh, dan persentase jumlah penduduk berdistribusi normal.

3.2 Uji Multikolinearitas

Tabel 3.3. Uji Multikolinearitas

Coefficients ^a		
Model	Collinearity Statistics	
	Tolerance	VIF
ln_jumlah_kasus	.709	1.410
Persentase Jumlah Penderita Sembuh	.892	1.122
ln_jumlah_penduduk	.698	1.433

Pada **Tabel 3.3** dapat dilihat sebuah output yang dihasilkan untuk menguji apakah terdapat multikolinearitas dalam analisis. Ketika suatu nilai toleransi (tolerance) lebih besar dari 0.10 maka artinya tidak terjadi multikolinieritas terhadap data yang diuji. Dapat dilihat bahwa nilai toleransi ketiga variabel masing-masing sebesar 0.709, 0.892, dan 0.698 maka lebih besar dari 0.10 sehingga bisa diduga bahwa antar variabel tidak terjadi multikolinieritas.

3.3 Cluster K-Means

Tabel 3.4. Rata-rata Cluster

```
> kmeans
K-means clustering with 4 clusters of sizes 12, 7, 12, 3

Cluster means:
  ln_pj_kasus Jp_Sembuh ln_j_penduduk
1 -2.595380  76.85269   0.8740596
2 -2.467769  61.89778   0.3329984
3 -2.568458  86.78179   0.4890626
4 -1.795031  51.23976  -1.5135524
```

Pada **Tabel 3.4** hasil pertama ditampilkan yaitu anggota kelompoknya dimana terdapat 4 kelompok (*Cluster*) dengan kelompok 1 memiliki 12 anggota, kelompok 2 memiliki 7 anggota, kelompok 3 memiliki 12 anggota, dan kelompok 4 memiliki 3 anggota.

Nilai *cluster means* yaitu nilai rata-rata variabel pada tiap kelompok yaitu kelompok 1 sampai 4. Pada nilai *cluster means* pertama menunjukkan jumlah rata-rata penderita kasus TB Paru terendah dibandingkan dengan *cluster* lainnya yaitu sebesar -2.595380, dan jumlah penderita yang sembuh tertinggi kedua dengan nilai rata-rata sebesar 76.85269, dengan rata-rata jumlah penduduk tertinggi sebesar 0.8740596.

Untuk kelompok kedua yaitu dengan rata-rata penderita kasus TB Paru kedua tertinggi yaitu sebesar -2.467769, dan jumlah penderita yang sembuh terendah kedua dengan rata-rata penduduk terendah kedua dengan masing-masing nilai 61.89778 dan 0.3329984.

Kelompok ketiga pengelompokkan dengan jumlah penderita TB Paru terendah kedua dan jumlah penderita yang sembuh tertinggi pertama dengan rata-rata jumlah penduduk tertinggi ke dua dibandingkan dengan cluster pertama dari masing-masing nilai sebesar 2.568458, 86.78179 dan 0.4890626.

Pada kelompok terakhir didapat pengelompokan dengan rata-rata jumlah penderita TB Paru tertinggi sebesar -1.795031 dan rata-rata jumlah penderita yang sembuh terendah sebesar 51.23976 dengan rata-rata jumlah penduduk terkecil sebesar -1.5135524.

Kelompok yang dihasilkan memiliki perbedaan dilihat dari *cluster means* yang dihasilkan, yaitu kelompok 1 sebagai kelompok dengan rata-rata tertinggi yang disusul dengan kelompok 3, lalu kelompok 2 dan terendah ada pada kelompok 4. Berdasarkan hal tersebut dapat disimpulkan bahwa kelompok 4 berhak dijadikan provinsi-provinsi yang menjadi prioritas ketika dilakukan program penanggulangan oleh pemerintah.

Tabel 3.5 Hasil Pengelompokkan dan Karakteristik kelompok

Kelompok	Anggota	Karakteristik
1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan, Kepulauan Riau, DKI Jakarta, Jawa Tengah, DI Yogyakarta, Nusa Tenggara Barat, Kalimantan Barat, Sulawesi Tenggara	Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terbanyak kedua di Indonesia, presentase jumlah kasus yang paling sedikit dan presentase jumlah penduduk yang paling banyak.
2	Jambi, Lampung, Jawa Barat, Kalimantan Tengah, Gorontalo, Maluku, Maluku Utara	Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terkecil kedua di Indonesia, presentase jumlah kasus yang lumayan banyak, dan presentase jumlah penduduk yang lumayan sedikit.
3	Bengkulu, Kep. Bangka Belitung, Jawa Timur, Banten, Bali, Nusa Tenggara Timur, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Barat.	Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terbesar di Indonesia, presentase jumlah kasus yang kecil, dan presentase jumlah penduduk yang banyak.
4	Kalimantan Utara, Papua Barat, Papua	Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terkecil di Indonesia, presentase jumlah kasus yang banyak, dan presentase jumlah penduduk yang kecil yaitu dibawah 2%.

Dari **Tabel 3.5** didapat provinsi-provinsi yang termasuk pada cluster 1, 2, 3 dan 4. Pada *cluster* 4 yang berhak dijadikan provinsi-provinsi yang menjadi prioritas ketika dilakukan sistem pencegahan oleh pemerintah yaitu provinsi Kalimantan Utara, Papua Barat, dan Papua. Kemudian disusul oleh kelompok 2 dengan nilai rata-rata tertinggi kedua dengan provinsi Jambi, Lampung, Jawa Barat, Kalimantan Tengah, Gorontalo, Maluku dan Maluku Utara. Dan provinsi-provinsi yang tidak menjadi prioritas ada pada kelompok 3 dan 1.

4. SIMPULAN

Dari berbagai hal yang telah dilakukan oleh peneliti, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Pengelompokkan Persentase Jumlah Kasus, Persentase Penderita Sembuh, Persentase Jumlah Penduduk berdasarkan Provinsi Indonesia tahun 2016 yaitu provinsi-provinsi yang berhak dijadikan prioritas ketika dilakukan sistem pencegahan oleh pemerintah adalah *cluster* 4 yaitu terdiri dari provinsi Kalimantan Utara, Papua Barat, dan Papua. Kemudian *cluster* 2 dengan nilai rata-rata tertinggi kedua yaitu Jambi, Lampung, Jawa Barat, Kalimantan Tengah, Gorontalo, Maluku dan Maluku Utara, dan provinsi-provinsi yang tidak menjadi prioritas ada pada *cluster* 3 dan 1.

2. Karakteristik sebaran Persentase Jumlah Kasus, Persentase Penderita Sembuh, Persentase Jumlah Penduduk dan Provinsi di Indonesia tahun 2016 menurut hasil pengelompokan setiap *clusternya* yaitu :
 - a. *Cluster 1* :Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terbanyak kedua di Indonesia, presentase jumlah kasus yang paling sedikit dan presentase jumlah penduduk yang paling banyak.
 - b. *Cluster 2* :Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terkecil kedua di Indonesia, presentase jumlah kasus yang lumayan banyak , dan presentase jumlah penduduk yang lumayan sedikit.
 - c. *Cluster 3* :Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terbesar di Indonesia, presentase jumlah kasus yang kecil, dan presentase jumlah penduduk yang banyak.
 - d. *Cluster 4*:Provinsi-provinsi dengan presentase kesembuhan penderita TB Paru terkecil di Indonesia, presentase jumlah kasus yang banyak , dan presentase jumlah penduduk yang kecil yaitu dibawah 2%.

5. DAFTAR PUSTAKA

- Agrawal, A. & Gupta, H., 2013. *Global K-Means (GKM) Clustering Algorithm: A Survey*. International Journal of Computer Applications, LIX(2), pp.20-24.
- Bangoria, B., Mankad, N. & Pambhar, V., 2013. *A Survey on Efficient Enhanced K-Means Clustering Algorithm*. International Journal for Scientific Research & Development, I(9), pp.1698-700.
- Budi, Purbayu Santosa dan Ashari. 2005. *Analisis Statistik dengan Microsoft Excel & SPSS*. Yogyakarta : Andi Offset.
- Chaturved, E.N. & Rajavat, E.A., 2013. *An Improvement in K-mean Clustering Algorithm Using Better Time and Accuracy*. International Journal of Programming Languages and Applications, III(4), pp.13-19.
- Depkes RI. 2017. *Profil Kesehatan Indonesia Tahun 2016*. www.depkes.go.id
- Eltibi, M.F. & Ashour, W.M., 2011. *Initializing K-Means Clustering Algorithm using Statistical Information. K-means clustering algorithm is one of the best known*, XXIX(7), pp.51-55.
- Gujarati, Damodar (1995). *Basic Econometrics*. (3rd edition) New York: McGraw Hill, Inc.
- Hastuti, Tiara, dkk (2015). “*Analisis Spasial, Korelasi dan Tren Kasus TB paru BTA positif menggunakan Web sistem Informasi Geografis di kota Kendari tahun 2013-2015*” vol 1, No 3 (2016), JIMKESMAS Jurnal Ilmiah Mahasiswa Kesehatan Masyarakat, Fakultas Kesehatan Masyarakat Universitas Halu Oleo.
- HUNG, C.M., WU, J., CHANG, J.H. & YANG, D.L., 2005. *An Efficient k-Means Clustering Algorithm Using Simple Partitioning*. JOURNAL OF INFORMATION SCIENCE AND ENGINEERING, XXI(1), pp.1157-77.
- Ifadah, Ana. 2011. *Analisis Metode Principal Component Analysis (Komponen Utama) Dan Regresi Ridge dalam Mengatasi Dampak Multikolinearitas Dalam Analisis Regresi Linear Berganda*. Skripsi. Jurusan Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam. Universitas Negeri Semarang.

- Madhulatha, T. S. (2012). *An Overview on Clustering Methods*. IOSR Journal of Engineering, 2(4).
- Santosa , P.B., dan Ashari. (2005). *Analisis Statistik Dengan Microsoft Excel dan SPSS*. Yogyakarta: ANDI.
- Saranya & Punithavalli, 2011. *An Efficient Centroid Selection Algorithm for K-means Clustering*. International Journal of Management, IT and Engineering, I(3), pp.130-40.