

Klastering Penyakit Tuberkulosis Menggunakan Algoritma K-Means (Studi Kasus Puskesmas Cisarua)

Reihan Salwa Fadhila Irawan¹, Dini Rohmayani²

^{1,2} Program Studi Teknik Informatika - Politeknik TEDC Bandung

Jl. Politeknik-Pesantren KM2 Cibabat Cimahi Utara – Cimahi Jawa Barat – Indonesia

salwaforster@gmail.com, dinirohmayani@poltektedc.ac.id

Abstrak— Tuberkulosis (TB) ialah infeksi menular yang diakibatkan oleh bakteri *Mycobacterium tuberculosis*, yang umumnya menyerang paru-paru serta bisa memengaruhi organ lain. Meskipun pengobatan dengan antibiotik telah tersedia, tantangan seperti resistensi obat dan kurangnya akses terhadap layanan kesehatan masih menghambat upaya pengendalian TB. TBC sangat berpengaruh dalam lingkungan maka melakukan skrining kesetiap daerah untuk mengetahui banyak atau tidaknya warga yang terkena penyakit TBC. Dimana di Puskesmas Cisarua terdapat beberapa daerah yang terkena penyakit TBC. Penyakit TBC juga bisa di menjadi tolak ukur seberapa baik tingkat pelayanan kesehatan di suatu daerah dikarenakan TBC juga masuk kedalam program pemerintah. Merujuk pada fenomena yang ada, penulis terdorong dalam melangsungkan *clustering* terhadap data penyakit TBC di Puskesmas Cisarua. Adapun metode yang digunakan untuk mengelompokkan data tersebut yakni algoritma *k-means*. Merujuk pada temuan uji, didapatkan bahwa nilai *k* yang optimal ialah $k=3$. Dengan demikian untuk pengelompokkan dataset yang di gunakan menjadi 3 kelompok *cluster*, dimana *cluster_0* sejumlah 245 data, *cluster_1* sejumlah 397 data dan *cluster_2* sejumlah 250 data. Dimana pada *cluster_1* di dominasi oleh RW 11 dengan dengan dominasi penyakit di dominasi oleh penyakit TBC dan banyak di derita oleh pasien berjenis kelamin laki-laki, penyebaran berdasarkan umur di dominasi oleh kategori dewasa (19-59). Sedangkan pada *cluster_0* banyak pasien yang terobati untuk penyakit tuberkulosis data didominasi oleh RW 14 banyak menyerang pasien berjenis kelamin Perempuan dan laki-laki. Sedangkan untuk penyebaran berdasarkan kategori umur didominasi oleh kategori dewasa. Algoritman *k-means* mendapatkan nilai *davies bouldin* sebesar -0.906. Temuan studi ini diproyeksikan mampu dijadikan sumber acuan bagi Puskesmas Cisarua untuk mengimplementasikan promosi kesehatan di wilayahnya tersebut.

Kata Kunci— Klastering, Algoritma *K-Means*, Penyakit Tuberkulosis, Puskesmas Cisarua.

Abstract— Tuberculosis (TB) is an infectious infection caused by the bacteria *Mycobacterium tuberculosis*, which generally attacks the lungs and can affect other organs. Although treatment with antibiotics is available, challenges such as drug resistance and lack of access to health services still hamper TB control efforts. TB is very affected in the environment, so we carry out screening in each area to find out whether or not many residents are affected by TB disease. Where in the Cisarua Health Center there are several areas affected by TB disease. TB disease can also be a benchmark for how good the level of health services in an area is because TB is also included in

the government program. Referring to the existing phenomenon, the author was encouraged to carry out clustering of TB disease data at the Cisarua Health Center. The method used to group the data is the *k-means* algorithm. Referring to the test findings, it was found that the optimal *k* value is $k=3$. Thus, to group the dataset used into 3 cluster groups, where *cluster_0* has 245 data, *cluster_1* has 397 data and *cluster_2* has 250 data. Where in *cluster_1* it is dominated by RW 11 with the predominance of the disease being dominated by TB disease and mostly male patients suffering from it, the distribution based on age is dominated by the adult category (19-59). Meanwhile, in *cluster_0*, many patients were treated for tuberculosis, data dominated by RW 14, which mostly attacked female and male patients. Meanwhile, the distribution based on age categories is dominated by the adult category. The *k-means* algorithm obtained a *Davies Bouldin* value of -0.906. The findings of this study are projected to be able to be used as a reference source for the Cisarua Community Health Center to implement health promotion in its area.

Keywords— Clustering, *K-Means* Algorithm, Tuberculosis Disease, Cisarua Health Center.

I. PENDAHULUAN

Paru-paru orang yang menderita tuberkulosis disebut paru, dan bakteri *Mycobacterium tuberculosis* (*M. tuberculosis*) bertanggung jawab atas penyakit menular ini. Global Tuberculosis Report mengungkapkan kepada publik bahwa pada tahun 2016, setelah India, india menduduki peringkat kedua dalam prevalensi tuberkulosis (TB). Nuryani dan kawan-kawan menegaskan bahwa penyakit ini dapat muncul pada usia berapa pun, berpotensi tanpa gejala, atau dengan gejala yang parah dan gambaran klinis yang bervariasi [1]. Sebagian besar kuman TB sering ditemukan menginfeksi parenkim paru dan menyebabkan TB paru, namun bakteri ini juga memiliki kemampuan menginfeksi organ tubuh lainnya (TB ekstra paru) seperti pleura, kelenjar limfe, tulang, dan organ ekstra paru lainnya [2].

Pada tahun 2021, proyeksi jumlah kasus tuberkulosis (TB) di Indonesia adalah 969.000 atau setara dengan 354 kasus per 100.000 orang. Saat ini, kasus tuberkulosis terkait HIV (TB-HIV) diperkirakan berjumlah 22.000 setiap tahunnya, setara dengan 8,1 kasus per 100.000 orang. Angka kematian akibat tuberkulosis (TB) yang diantisipasi adalah 144.000 orang, setara dengan 52 kematian per 100.000 penduduk, sedangkan

kematian akibat TB-HIV berjumlah 6.500 orang, atau 2,4 kematian per 100.000 penduduk.

Analisis statistik kejadian tuberkulosis dari tahun 2000 hingga 2020 menunjukkan adanya tren penurunan baik jumlah kasus maupun angka kematian, namun penurunannya tidak signifikan. Meski demikian, dalam kurun waktu 2020-2021 terjadi peningkatan kasus dan kematian yang signifikan. Pada tahun 2021, angka kejadian tuberkulosis meningkat sebesar 18%, dengan jumlah absolut meningkat dari 819.000 kasus pada tahun 2020 menjadi 969.000 kasus pada tahun 2021. Angka kejadian per 100.000 orang meningkat dari 301 pada tahun 2020 menjadi 354 pada tahun 2021.

Puskesmas merupakan unit organisasi bersifat fungsional dan unit layanan yang bekerja secara profesional [3]. Untuk mengetahui daerah tingkat penyakit tuberkulosis tertinggi di Puskesmas Cisarua yang perlu ditangani maka dapat menggunakan teknik data mining.

Menurut Sulisty, angka kematian akibat tuberkulosis telah meningkat sebesar 55% secara absolut, meningkat dari 93.000 kematian pada tahun 2020 menjadi 144.000 pada tahun 2021. Angka kematian per 100.000 orang meningkat sebesar 52%, meningkat dari 34 kematian pada tahun 2020 menjadi 52 kematian pada tahun 2021 [4].

Dari permasalahan yang dipaparkan pada paragraf sebelumnya, diluncurkan suatu studi yang bertujuan dalam mengevaluasi rendah dan tingginya tingkat kasus penyakit tuberkulosis berdasarkan Rukun warga (RW), Rukun Tetangga (RT), jenis kelamin, umur, serta mengetahui pasien terobati dan tidak terobati di daerah Cisarua, dengan judul penelitian yaitu “Clustering Penyakit Tuberkulosis Menggunakan Algoritma K-Means (Studi Kasus Puskesmas Cisarua)”.

II. METODE PENELITIAN

Dari karya ilmiah Sahir, metode penelitian berupa rangkaian kegiatan yang menimbulkan hipotesis-hipotesis. Dengan dibantu persepsi penelitian, untuk nantinya data dapat diolah dan dianalisa yang akan membentuk suatu kesimpulan [5]. Teknik pengumpulan data dapat digunakan untuk mengidentifikasi wilayah di Puskesmas Cisarua yang memiliki tingkat kasus tuberkulosis tertinggi sehingga memerlukan pengobatan prioritas.

Data mining menurut Wardhani merupakan teknik pemrosesan data yang dirancang untuk mengungkap pola tersembunyi di dalam data untuk menghasilkan wawasan baru [6]. Teknik-teknik ini mencakup beberapa metode termasuk asosiasi, deskripsi, estimasi, klasifikasi, pengelompokan, dan prediksi. Untuk mendapatkan data diperlukan jumlah *cluster* awal sebagai masukan dan menghasilkan *cluster* akhir [7]. Penelitian ini memilih metode *clustering* sebagai pendekatan yang digunakan.

Metode *clustering* yang diterapkan pada data kasus tuberkulosis di Puskesmas Cisarua menunjukkan bahwa wilayah dengan angka kejadian tertinggi terkonsentrasi pada kelompok tertentu yang menunjukkan ciri-ciri serupa, termasuk peningkatan kepadatan penduduk dan terbatasnya akses terhadap layanan kesehatan. Kelompok ini ditetapkan sebagai target utama intervensi, sehingga memungkinkan

pengembangan langkah-langkah pengobatan yang ditargetkan, termasuk peningkatan fasilitas kesehatan, inisiatif kesadaran masyarakat, dan pelacakan kasus yang ketat. Hasilnya menunjukkan bahwa metode *clustering* berguna dalam menetapkan prioritas pengobatan berdasarkan pola data.

A. Pengumpulan Data

Data primer adalah informasi yang diperoleh langsung dari sumbernya dan berkaitan dengan penelitian yang dilakukan. Penelitian ini memperoleh data asli langsung dari Puskesmas Cisarua yang berjumlah 914 titik data.

Data sekunder mengacu pada informasi yang diperoleh secara tidak langsung dari entitas lain yang telah memproses data tersebut. Penelitian ini memanfaatkan data sekunder dari beberapa sumber antara lain jurnal, buku, dan artikel terkait Puskesmas Cisarua, tuberkulosis, data mining, metodologi *clustering*, algoritma K-Means, dan *software* RapidMiner.

Pengumpulan data juga dilakukan melalui observasi langsung ke lapangan. Penelitian ini meliputi observasi yang dilakukan di Puskesmas Cisarua di Kecamatan Cisarua, Kabupaten Bandung Barat, Jawa Barat, dengan fokus utama pada pengumpulan data terkait penyakit tuberkulosis.

Selain itu, studi pustaka juga menjadi bagian penting dalam pengumpulan data. Penulis melakukan kegiatan membaca, mencatat, dan menganalisis berbagai literatur yang relevan untuk memperoleh teori-teori pendukung terkait penyakit tuberkulosis dan penerapan data mining.

B. Tahapan K-Means

Proses ini umumnya dilakukan menggunakan algoritma dasar dengan langkah-langkah berikut [8] :

1. Menentukan jumlah kluster yang diinginkan
2. Menghasilkan centroid (titik pusat kluster) secara acak
3. Menghitung jarak centroid atau rata-rata dari data dalam setiap kluster

Untuk melakukan perhitungan jarak diantara kluster, *Euclidean Distance* merupakan algoritma yang sering digunakan, dengan rumus sebagai berikut:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dimana :

- x : Titik data pertama,
 - y : Titik data kedua,
 - n : Jumlah karakteristik (atribut) dalam terminologi data mining
 - d : *Euclidean Distance* yaitu jarak antara data pada (x,y) titik x dan titik y menggunakan kalkulasi matematika
4. Mengelompokkan setiap data ke *centroid* atau rata-rata terdekat dengan membandingkan jarak. Data akan dimasukkan ke dalam kluster yang memiliki jarak terdekat dari *centroid*.
 5. Menghitung rata-rata nilai dari data yang berada dalam kluster yang sama untuk menentukan *centroid* baru. Proses ini diulangi dengan menggunakan pusat kluster yang telah

diperbarui hingga mencapai kondisi konvergen atau hasil yang stabil.

Kembali ke langkah ketiga, jika masih terdapat data yang berpindah klaster, atau jika perubahan nilai centroid melebihi ambang batas (*threshold*) yang telah ditentukan, atau jika perubahan nilai pada fungsi objektif yang digunakan melampaui ambang batas tersebut.

C. Preparasi Data

Merujuk pada penjabaran dari Kementerian Kesehatan Republik Indonesia (Kemenkes) Usia dapat dikelompokkan dari usia 0-5 Tahun adalah Bayi (1), 5-9 Tahun Anak-Anak (2), 10-18 Tahun Remaja (3), 19-59 Tahun Dewasa (4), dan diatas 60 Tahun termasuk Lansia (6) [9].

Jenis kelamin merujuk pada perbedaan biologis antara perempuan (1) dan laki-laki (2), yang merupakan karakteristik bawaan sejak lahir dan bersifat permanen serta tidak dapat diubah derajatnya [10].

Rukun Warga merupakan istilah pembagian wilayah di bawah kelurahan yang merupakan alamat pasien penderita penyakit tuberkulosis. Ada 19 RW yang penulis akan teliti. Rukun Tetangga merupakan organisasi yang terdiri dari sejumlah Kepala Keluarga (KK) di setiap Desa atau Kelurahan, yang dikoordinasikan oleh seorang ketua.

Kategori terobati atau tidak terobati yang dikonversikan menjadi angka dimana terobati 3 dan tidak terobati 4. hasil inisialisasi dari tiap-tiap atribut yang menghasilkan 5 atribut agar terbaca oleh aplikasi *rapidminer*, seperti pada TABEL I berikut.

TABEL I
HASIL INISIALISASI DATA

No. Urut	Umur	Jenis Kelamin	RW	RT	Terobati/Tidak Terobati
1	3	2	3	3	3
2	4	2	8	2	3
3	4	1	4	2	3
4	4	2	11	1	3
5	4	2	15	3	3
6	1	2	8	1	3
7	4	2	5	5	3
8	1	2	7	2	3
9	4	2	11	5	3
10	4	2	2	2	3
.....
892	4	1	11	2	4

D. Hasil

Dataset pasien penyakit tuberkulosis dikelompokkan berdasarkan atribut, data dan atribut tersebut dikelolah menggunakan aplikasi *RapidMiner* berfungsi untuk menggunakan algoritma *clustering k-means* yang nantinya akan berpotensi untuk mendapatkan atau menghasilkan *centeroid* dan *cluster*. Hasilnya akan mengelompokkan data sesuai dengan kemiripan data.

Data yang awalnya berjumlah 914 data, disaring menjadi 892 data karena adanya beberapa data yang tidak lengkap untuk

pengolahan lebih lanjut, seperti yang tercantum pada Tabel I. Dalam penelitian ini, penulis hanya menggunakan 5 atribut dari total 20 atribut yang ada, yaitu umur, jenis kelamin, serta daerah seperti RT dan RW, serta status pengobatan (terobati dan tidak terobati). Kelima atribut ini dipilih karena relevansinya dengan kasus tuberkulosis, di mana faktor umur merupakan faktor pertama yang berpengaruh, mengingat insiden tertinggi tuberkulosis terjadi pada kelompok usia tertentu. Faktor kedua adalah jenis kelamin, karena penyakit ini lebih banyak menyerang laki-laki, yang umumnya memiliki kebiasaan merokok. Faktor ketiga adalah faktor lingkungan, terutama pada penderita tuberkulosis, karena kuman M. Tuberculosis dapat berkembang di rumah dengan kondisi gelap dan minim cahaya matahari.

E. Analisis

Sebagaimana yang telah penulis jelaskan dibab sebelumnya, bahwa penelitian ini merupakan *peng-clusteran* data penyakit tuberkulosis yang ada di Puskesmas Cisarua. Sebelumnya juga telah dilakukan pengumpulan data, melakukan observasi, mencari studi pustaka dari berbagai sumber, serta pengolahan data dan yang terakhir melakukan preparasi data. Pada penelitian ini penulis melakukan pengolahan data awal terdahulu yang didapat langsung dari Puskesmas Cisarua, setelah itu diolah data mana saja yang hendak dimanfaatkan pada studi ini dengan dilakukannya preparasi data. Setelah melakukan preparasi data menghasilkan 5 atribut dari 20 atribut serta data dari 222 menjadi 201 data yang didapat. Selanjutnya nanti akan dilakukannya proses *clustering* dengan algoritma *k-means* dilakukan menggunakan aplikasi *rapidminer*, hasil pengelompokkan data berdasarkan penyakit TBC memberikan gambaran yang signifikan mengenai persebaran penyakit di daerah Cisarua. Analisis ini dilakukan dengan melihat pola distribusi pasien yang dikelompokkan berdasarkan atribut seperti umur, jenis kelamin, dan wilayah seperti RT dan RW serta terobati dan tidak terobati.

III. HASIL DAN PEMBAHASAN

Dari proses penentuan kluster di dapatkan kluster empat dengan nilai *Davies Bouldin Index* (DBI) menyentuh kategori paling optimal. Maka dari itu, data dilakukan pengelompokkan pada penelitian ini menjadi dilakukan pembagian pada 3 kelompok yaitu kluster 0 dengan banyak jumlah data 245, pada kluster 1 berjumlah 397 data, kluster 2 berjumlah 250 dengan total 892 data. Pengelompokkan data tersebut dapat di lihat pada Gbr 1 berikut.

Cluster Model

```
Cluster 0: 245 items
Cluster 1: 397 items
Cluster 2: 250 items
Total number of items: 892
```

Gbr 1 Pengelompokkan Data Sesuai Cluster

Pada Gbr 2 menampilkan data yang telah di olah secara keseluruhan lengkap dengan jenis *cluster*-nya.

Row No.	No Urut	cluster	Usia	Jenis Kel...	RW	RT	Teknikal Td...
1	1	cluster_1	4	2	3	2	3
2	2	cluster_2	4	2	8	2	3
3	3	cluster_1	4	2	4	2	1
4	4	cluster_2	4	1	11	1	3
5	5	cluster_0	4	2	15	2	3
6	6	cluster_2	1	2	8	1	3
7	7	cluster_1	4	2	5	5	3
8	8	cluster_2	1	2	7	2	3
9	9	cluster_0	4	2	11	8	3
10	10	cluster_1	4	2	2	2	3

ExampleSet (892 examples, 2 special attributes, 5 regular attributes)

Gbr 2 Data hasil setiap cluster menggunakan *exampleset* (apply Model)

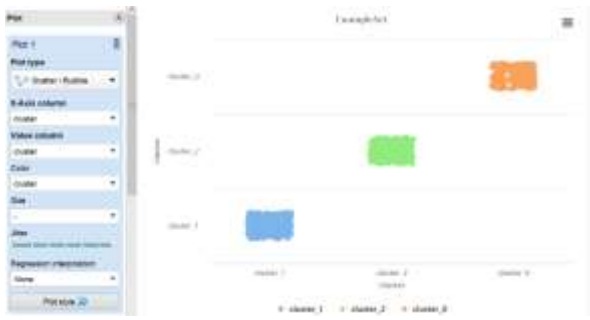
Data penyakit TBC yang ada pada gambar diatas pada *cluster* 0 tersusun dari 245 data. Data penyakit TBC yang termasuk pada kategori *cluster* 1 tersusun dari 397 data. Data penyakit TBC yang termasuk kedalam *cluster* 2 terdiri dari 250 data. Setelah melakukan pengujian maka ada hasil berdasarkan rata-rata *within centroid distance* dapat di lihat pada Gbr 3.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -5.730
Avg. within centroid distance_cluster_0: -6.882
Avg. within centroid distance_cluster_1: -5.454
Avg. within centroid distance_cluster_2: -5.040
Davies Bouldin: -0.906
```

Gbr 3 Hasil *PerformanceVector* Dengan Nilai DBI

Sedangkan untuk grafik plot data *View* berdasarkan *cluster* yang telah di tentukan di awal yaitu berjumlah 3 *cluster* dimana ada *cluster* 0, lalu *cluster* 1 dan *cluster* 2 maka grafik penyebaran data nya bisa diamati di Gbr 4.



Gbr 4 Plot Data Penyebaran Data Berdasrkan Cluster

Melalui pemanfaatan algoritma *k-medoids clustering*, diperoleh temuan yang menunjukkan terbentuknya 5 klaster, sesuai dengan penentuan nilai k. Jumlah data pada masing-masing klaster adalah sebagai berikut: klaster 0 berisi 371 data, klaster 1 berisi 150 data, klaster 2 berisi 54 data, klaster 3 berisi 166 data, dan klaster 4 berisi 151 data, dengan total keseluruhan data sebanyak 892. Hasil pengelompokkan data bisa di lihat pada Gbr 5.

Cluster Model

```
cluster 0: 371 items
cluster 1: 150 items
cluster 2: 54 items
cluster 3: 166 items
cluster 4: 151 items
Total number of items: 892
```

Gbr 5 Hasil Pengelompokkan Data Dengan Algoritma *K-Medoids*

Sedangkan untuk grafik Plot Data *View* berdasarkan *cluster* yang telah di tentukan di awal yaitu berjumlah 4 *cluster* dimana ada *cluster* 0, lalu *cluster* 1, juga ada *cluster* 2, kemudian *cluster* 3, dan *cluster* 4 maka grafik penyebaran data nya bisa diamati di Gbr 6.



Gbr 6 Plot Data *View* Penyebaran Data Algoritma *K-Medoids*

Algoritma *k-means* dan *k-medoids* menciptakan klaster yang berbeda serta memiliki nilai yang berbeda pula. Dalam pengujian kali ini, *algoritma k-means* menunjukkan nilai DBI yang lebih optimal dibandingkan dengan algoritma *k-medoids*. Hasil perbandingan tersebut dapat di lihat pada TABLE II.

TABEL II
HASIL PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS*

Algoritma	Nilai DBI	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
K-Means	-0.906	245 items	397 items	250 items	-	-
K-Medoids	-1.762	371 items	150 items	54 items	166 items	151 items

Pada hasil pengujian yang di lakukan peneliti terdapat perbedaan beberapa hal dalam proses pengolahan data menggunakan data mining. Yang pertama adalah perbedaan nilai *Davies Boulding Index* (DBI) yang di dihasilkan oleh setiap algoritma dengan pengaturan parameters yang sama. Pada pengujian nilai DBI algoritma *k-means* menghasilkan nilai yang lebih optimal dari pada algoritma *k-medoids*, dimana skor DBI dipakai sebagai acuan jumlah kluster yang memperlihatkan klasterisasi terbaik atau paling relevan meskipun di algoritma *k-means* menghasilkan 3 *cluster*

sedangkan k-medoids menghasilkan 4 *cluster*. Adapun hasil perbandingan nilai *validasi Davies Boulding Index* (DBI) antara *k-means* dan *k-medoids* pada dtufi ini bisa diamati di TABEL III.

TABEL III
HASIL PERBANDINGAN NILAI *DAVIES BOULDIN INDEX* (DBI)

Pengujian Nilai K	K-Means	K-Medoids
k= 2	-0.686	-3.356
k=3	-0.906	-2.166
k=4	-1.046	-1.846
k=5	-1.265	-1.762

Berdasarkan hasil TABEL III diatas dapat di simpulkan bahwa nilai DBI terbaik di hasilkan oleh algoritma k-means dengan k=3 dengan nilai yang hampir mendekati 1 yaitu -0.906 yang di jadikan acuan pengelompokkan data pada penelitian kali ini.

Selanjutnya terdapat perbedaan juga didalam pengelompokkan data set oleh kedua algoritma tersebut ketika pengelompokkan data dengan kelompok yang berbeda. Algoritma *k-means* menghasilkan 3 klaster, dengan distribusi jumlah item sebagai berikut: klaster 0 memiliki total 397 item, klaster 1 memiliki total 245 item, dan klaster 2 memiliki total 250 item. Sementara itu, algoritma k-medoids menghasilkan 4 klaster, dengan distribusi jumlah item sebagai berikut: klaster 0 berjumlah 371 item, klaster 1 berjumlah 150 item, klaster 2 berjumlah 54 item, klaster 3 berjumlah 166 item, dan klaster 4 berjumlah 151 item. Perbandingan pengelompokkan data berdasarkan cluster antara algoritma k-means dan k-medoids dapat di lihat pada TABEL IV.

TABEL IV
HASIL PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS*

Cluster	K-Means	K-Medoids
<i>Cluster 0</i>	245	371
<i>Cluster 1</i>	397	150
<i>Cluster 2</i>	250	54
<i>Cluster 3</i>	-	166
<i>Cluster 4</i>	-	151

Dari beberapa perbedaan di atas dapat di simpulkan bahwa algoritma *k-means* lebih baik dalam pengelompokkan data penyakit TBC di Puskesmas Cisarua dari pada algoritma *k-medoids*.

Berdasarkan temuan uji melalui algoritma k-means, diperoleh klaster optimal dengan k=3. Oleh karena itu, dataset yang digunakan dikelompokkan menjadi 3 klaster, yaitu klaster_0 dengan 245 data, klaster_1 dengan 397 data, dan klaster_2 dengan 250 data. Terkait hasil data klaster k-means, terdapat beberapa data yang dominan di setiap klaster tersebut.

Sedangkan untuk hasil analisis pendukung lainnya, jumlah penyebaran kasus pasien terobati dan tidak terobati di wilayah Puskesmas Cisarua berdasarkan cluster dapat di lihat pada TABEL V. Berikut adalah tabel jumlah pasien terobati dan tidak terobati penyakit TBC di wilayah Puskesmas Cisarua.

TABEL V
JUMLAH KASUS PENYAKIT TBC TEROBATI DAN TIDAK TEROBATI

Kategori	Cluster_0	Cluster_1	Cluster_2
Terobati	88	56	52
Tidak Terobati	157	341	198

Dari data diatas dapat diambil beberapa kesimpulan dimana pada ketiga cluster tersebut didominasi oleh kasus tidak terobati oleh penyakit tuberkulosis. Hal ini menunjukkan bahwa pada RW 11 termasuk cluster_1 di Puskesmas Cisarua kasus tertinggi penyakit TBC tidak terobati. Untuk jenis kelamin yang rentan terkena penyakit TBC pada data yang peneliti lihat, bahwa laki-laki lebih rentan terkena penyakit TBC. Dimana pada tiga cluster yang ada, data penyakit TBC di dominasi oleh pasien berjenis kelamin laki-laki ini didasari oleh pola dan kebiasaan yang dilakukan seperti merokok, lingkungan tempat kerja, serta penyakit menular.

Untuk penyebaran berdasar kategori umur, pada *cluster_0* didominasi oleh kategori usia Dewasa (19-59 tahun) dimana tercatat ada 132, kedalam kategori umur tersebut pada *cluster_1* lebih didominasi dewasa usia (19-59) tercatat ada 223, sedangkan untuk *cluster_2* tercatat 142 didominasi oleh dewasa (19-59 tahun). Hal ini menjelaskan bahwa pada RW 11 banyak sekali anak dewasa umur kisaran (19-59 tahun) yang tidak terobati. Berikut adalah TABEL VI jumlah kasus penyakit TBC dengan kategori umur.

TABEL VI
JUMLAH KASUS PENYAKIT TBC TEROBATI DAN TIDAK TEROBATI

Umur	Kategori	Cluster_0	Cluster_1	Cluster_2
0-5	Bayi dan balita	23	41	22
5-9	Anak-anak	11	26	17
10-18	Remaja	39	41	32
19-59	Dewasa	132	223	142
>60 tahun	Lansia	40	66	37

Dari beberapa hasil analisis berkaitan dengan pengelompokkan data penyakit TBC memakai algoritma dari *k-means clustering* diatas maka peneliti dapat menyimpulkan beberapa saran untuk menekan angka penyebaran penyakit TBC di daerah Puskesmas Cisarua. Berikut adalah beberapa upaya yang dapat dilakukan :

1. Melakukan promosi Kesehatan atau penyuluhan kesehatan tentang pencegahan dan pengendalian penyakit TBC khususnya penyakit TBC yang banyak terjadi. Dalam hal ini berdasarkan hasil pengelompokkan data dari *cluster 0*, *cluster 1*, *cluster 2* didominasi umur dewasa (19-59 tahun) yang terjangkit, maka dari itu adakan penyuluhan terhadap anak dewasa.
2. Melakukan penyuluhan kesehatan dengan memfokuskan kedaerah sesuai dengan hasil pengelompokkan menggunakan *k-means clustering* yaitu pada daerah dengan kasus tertinggi penyakit TBC yang termasuk kedalam *cluster_1* yaitu RW 11.
3. Melakukan skrining kesetiap rumah warga agar terdeteksi sejak awal, hal ini bisa mencegah atau mengurangi kasus penyakit TBC. Pada kasus TBC berikan obat rutin kepada

pasien yang sudah terkena ini bisa mengurangi penularan terhadap banyak orang.

IV. PENUTUP

Berdasarkan hasil pengujian serta hasil pembahasan yang dilakukan oleh penulis tentang pengelompokan data untuk mengetahui penyakit TBC menggunakan *K-Means clustering*, dengan demikian bisa disimpulkan seperti berikut :

- 1) Pada proses *clustering* melakukan inisialisasi terdahulu agar pada saat dimasukkan ke dalam *rapidminer* bisa terbaca pada pengujian ini melakukan import data selanjutnya menggunakan *apply model* dan *performance* untuk melihat nilai K, pada *performance* parameternya diubah menggunakan *davies bouldin* agar pada saat eksperimen menghasilkan K terbaik selanjutnya akan mengetahui *cluster* mana yang terbaik yang akan dijadikan acuan untuk menentukan *cluster*.
- 2) Merujuk pada temuan uji melalui algoritma k-means, didapatkan total klaster optimal yakni k=3. Maka, dataset yang digunakan dikelompokkan ke dalam 3 klaster, dengan distribusi data sebagai berikut: klaster_0 yakni sejumlah 245 data, klaster_1 yakni sejumlah 397 data, dan klaster_2 yakni sejumlah 250 data. Hasil pengelompokan data dengan k-means menunjukkan adanya dominasi data tertentu di setiap klaster. Dimana pada *cluster 1* di dominasi oleh RW 11 dengan dengan dominasi penyakit di dominasi oleh RW 11 di daerah Jambudipa dengan penyakit TBC dan banyak di derita oleh pasien berjenis kelamin perempuan, penyebaran berdasarkan umur di dominasi oleh kategori umur Dewasa (19-59). Sedangkan pada *cluster_0* data didominasi trobati sebanyak 88 banyak menyerang pasien berjenis kelamin laki-laki dan perempuanimbang.

Pada penelitian kali ini mungkin terdapat banyak kekurangan dalam pelaporan maupun hasil pengujian pada

penelitian kali ini. Maka, dalam pengembangan secara komprehensif, di perlukan saran seperti berikut :

- 1) Sebaiknya pada penelitian selanjutnya diharapkan dapat menggunakan algoritma lainnya seperti algoritma *Fuzzy C-Means* atau parameters ang berbeda dalam memperoleh temuan yang lebih optimal lagi.
- 2) Pada studi berikutnya, diproyeksikan untuk menerapkan data yang lebih lengkap lagi seperti catatan anamnesa agar penyakit lebih lengkap lagi serta dengan cakupan data yang lebih luas seperti data pada dinas Kesehatan kabupaten kota atau dinas kesehatan Kesehatan provinsi untuk penyempurnaan hasil pengujian *clustering*.

REFERENSI

- [1] Nuryani, S., Nursilmi, D. L., & Dina, "Tuberculosis Di Rumah Sakit Umum X Kota Bandung". 1(November), 1601–1607, 2021.
- [2] Putranto, T. A., "Pedoman Nasional Pelayanan Kedokteran Tentang Tata Laksana Tuberkulosis (S. Prof.Dr.dr.Sudigdo Sastroasmoro (ed.); Cetakan I)", 2020.
- [3] Kementerian Kesehatan RI. Moeloek, N. F., "Permenkes nomor 43 tahun 2019 tentang puskesmas", 2019.
- [4] Sulisty, "Laporan Program Penanggulangan Tuberkulosis Tahun 2022". Kemenkes RI, 1–156, 2023.
- [5] Sahir, S., "METODOLOGI PENELITIAN (M. s. Dr.Ir. Koryati (ed.); Cetakan I)". KBM Indonesia, 2021.
- [6] Wardhani, A. K., "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan". Jurnal Transformatika, 14(1), 30, 2016.
- [7] Chandra, M. D., Irawan, E., Saragih, I. S., Windarto, A. P., & Suhendro, D., "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi". BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer, 2(1), 30–38, 2021.
- [8] N. S. Rahayu, Data Mining Dan Penerapan Algoritma, Cetakan 1, 2021.
- [9] Kemenkes, "Pengelompokan Usia Ayo Sehat Kementerian Kesehatan."
- [10] J. P. M. Tangkudung, "Proses Adaptasi Menurut Jenis Kelamin dalam Menunjang Studi Mahasiswa Fisip Universitas Sam Ratulangi," J. "Acta Diurna, vol. 3, no. 4, pp. 1–11, 2014.