
IMPLEMENTASI K-MEANS TERHADAP PENYEBARAN PENYAKIT TBC DI RIAU MENGGUNAKAN RAPID MINER

Dafwen Toresa¹

¹Teknik Informatika, Universitas Lancang Kuning

Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru

¹ dafwen@unilak.ac.id

Abstrak

Tuberkolusis merupakan penyakit paru-paru disebabkan oleh Mycobacterium tuberculosis. TBC dapat menimbulkan gejala berupa batuk lama, Tuberkulosis diderita orang kebanyakan merupakan infeksi TBC, yaitu terdapat bakteri TBC yang pasif. Dengan menggunakan algoritma kmeans dan aplikasi rapid miner memudahkan dalam melakukan klasterisasi wilayah sehingga tindakan dan penanganan untuk memberantas penyakit TBC dilakukan dengan tepat sasaran dan efisien. K-means adalah algoritma klasterisasi yang bekerja berdasarkan bagian dari clustering. Sehingga dapat menghasilkan iterasi clustering pada data jumlah penderita TBC di Riau. Hasil dari pengujian yang dikerjakan pada penelitian ini, iterasi clustering pada data jumlah penderita TBC di Riau tersebut dapat diolah menggunakan aplikasi bantuan Rapid miner yang digunakan untuk menentukan dari nilai centroid pada 3 cluster yaitu cluster Rendah (C1), cluster sedang (C2) dan cluster Tinggi (C3). Dan dari 12 wilayah yang masuk kedalam cluster tingkat Rendah untuk penderita kasus TBC 3 wilayah, 7 wilayah cluster tingkat sedang dan 2 provinsi lainnya termasuk cluster tingkat Tinggi

Kata kunci : TBC, Riau, k-means, cluster, Kesehatan.

Abstract

Tuberculosis is a lung disease caused by Mycobacterium tuberculosis. TB can cause symptoms such as a long cough. Tuberculosis suffered by large people is a TB infection, which contains passive TB bacteria. By using the K-means algorithm and fast mining applications help in clustering the region, so that helping and eradicating TB disease is carried out as desired and efficient. K-means is a clustering algorithm that works based on parts of a grouping. So that it can produce clustering iteration on the number of TB patients in Riau. The results of the tests carried out in this study, grouping iterations in the data on the number of TB patients in Riau can be processed using the Rapid Miner application that is used to determine the centroid value of 3 clusters, namely the Low cluster (C1), the Medium cluster (C2) and the cluster Height (C3). And of the 12 regions included in the low-level cluster for patients with TB cases 3 regions, seven regions with middle level and two other provinces included with the level of level

Keywords: TBC, Riau, k-means, clustering, Health.

I. PENDAHULUAN

TBC (Tuberkulosis) biasa diketahui dengan istilah TB merupakan penyakit yang menyerang paru-paru yang disebabkan oleh kuman *Mycobacterium tuberculosis*. TB dapat menyebabkan gejala penyakit berupa batuk yang berlangsung sangat lama (lebih dari 3 minggu), mengandung lender dahak, dan terkadang bagi beberapa penderita lainnya dapat mengeluarkan darah. Penyakit TB tidak hanya menyerang paru-paru, tetapi juga bisa menyerang bagian lainnya seperti usus, tulang, atau kelenjar pada tubuh. TB ini dapat metular dari cairan air liur yang keluar dari penderita TB, disarankan ketika berbicara atau batuk, dan bersin bagi penderita TB harap menutup mulut atau menggunakan masker. Penyakit ini lebih mudah menular kepada orang yang memiliki kekebalan tubuh yang rendah, misalnya penderita HIV.

Tuberkulosis merupakan infeksi TBC, yaitu merupakan bakteri TB yang tidak aktif secara klinis[1]. kuman TB mulai aktif dan menimbulkan reaksi ketika pasien menunjukkan tanda setelah periode waktu tertentu berlalu, dapat beberapa minggu bahkan beberapa tahun lamanya, tergantung dari keadaan kesehatan dan daya tahan penderita.

Di Indonesia, Pemerintah berupaya melakukan akselerasi dalam upaya pengurangan jumlah penderita TB pada 2030. percepatan dilakukan dengan cara pencegahan, diagnosis, memberikan pengobatan secara cermat, dan memberika layanan kesehatan bagi seluruh penderita TB dengan baik, serta berusaha dalam meningkatkan pendanaan program TBC yang berkesinambungan dan mandiri. Dikarenakan 5000 jiwa meninggal setiap harinya karena TBC. Penyakit TBC ini memiliki keterkaitan yang sangat erat dengan kondisi kesehatan masyarakat yang sangat buruk dan jumlah angka kemiskinan yang masih tinggi. Saat ini masalah penyakit

TBC semakin bertambah dengan kemunculan *multi-drug resistant TBC*.

Pemeriksaan tahap awal Tuberkulosis yang diketahui lebih dini menjadi hal yang sangat penting dalam mencegah penularan dan tindakan pencegahan dengan diobati lebih lanjut, pemeriksaan dan pengobatan TBC, menurut data dari Dinas Kesehatan Provinsi Riau menyatakan, sedikitnya 31.000 masyarakat di Provinsi Riau menderita TB. Sebanyak kurang lebih 20.000 belum dapat menjalani pengobatan sehingga besarnya potensi penyebaran penyakit ke masyarakat lain sangat tinggi [2]. dikarenakan TBC masih penyakit satu dari 10 penyebab angka kematian dan penyebab utama besarnya angka kematian di Indonesia. sedangkan untuk di seluruh dunia dapat diperkirakan kurang lebih 10 juta kasus TB baru setara 133 kasus per 100.000 penduduk [3] di dunia.

Sehingga sangat di perlukan pemetaan serta pengelompokan penyebaran dan penderita penyakit TBC di Riau. Karena merupakan salah satu provinsi yang memiliki angka penderita penyakit TBC cukup tinggi. Diperlukan sebuah pemodelan berbentuk klaster dan di kelompokkan berdasarkan jumlah penderita pada masing-masing wilayah dapat memudahkan pemerintah dalam menangani dan menindaklanjuti penyebaran TBC ini. Sehingga dengan dilakukan klasterisasi wilayah dan jumlah penderita. Pemerintah mengetahui dimana kelompok wilayah yang masuk dalam kategori tinggi maupun rendah .

Dengan menggunakan algoritma kmeans dan aplikasi rapid miner memudahkan dalam melakukan klasterisasi wilayah sehingga tindakan dan penanganan untuk memberantas penyakit TBC dilakukan dengan tepat sasaran dan efisien.

II. TINJUAN PUSTAKA

Data mining atau dikenal juga sebagai knowledge discovery in database (KDD) dapat digunakan secara silih berganti dengan menjelaskan proses penggalian sebuah informasi yang tersembunyi [4]. Berdasarkan riset pada penelitian prediksi kedatangan jumlah wisatawan ke kota Batam menggunakan teknik KDD, dapat menghasilkan angka yang cukup akurat dalam memprediksi berapa jumlah wisatawan yang datang ke kota Batam [5].

Clustering merupakan salah satu metode untuk pengelompokan data. Berdasarkan penelitian Tan, clustering mampu melakukan kegiatan proses untuk mengelompokkan data kedalam banyak cluster atau kelompok sehingga data pada satu cluster memiliki tingkat kemiripan yang sejenis dan data antar cluster juga memiliki kemiripan yang minimum[6].

K-means merupakan teknik dalam pengelompokan yang dapat bekerja berdasarkan pembagian cluster. Pada dasarnya prinsip kinerja dari kegiatan pengelompokan hierarchical clustering dilakukan secara setahap demi setahap[7]. Dan pada setiap iterasi dari pengelompokan hierarchical clustering hanya ada sekali pemilihan untuk penggabungan suatu item pada item lainnya. Tahapan dalam melakukan Clustering dengan metode K-Means adalah sebagai berikut[8]:

1. Pilih jumlah cluster k.
2. Lakukan inisialisasi k pusat cluster dengan berbagai cara. Tahap yang sering digunakan dengan cara random. pusat cluster diberi nilai awal dengan nilai random.
3. Pindahkan semua data ke cluster terdekat. Kedekatan dua objek ditentukan pada jarak kedua objek itu sendiri. cara menghitung jarak semua data ke titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut [9]:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

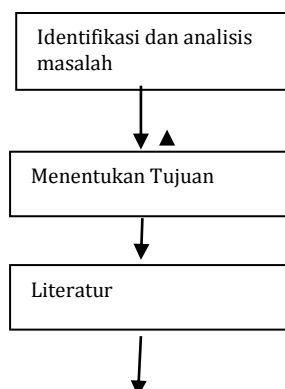
Keterangan:

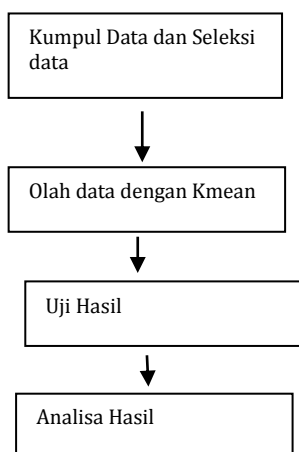
$D(i,j)$ = Jarak data ke i ke pusat cluster j
 X_{ki} = Data ke i atribut data ke k
 X_{kj} = Titik pusat ke j pada atribut ke k

4. Hitung terus jarak pusat cluster dengan anggota cluster yang ada sekarang. Pusat cluster merupakan rata-rata dari semua data/objek dalam cluster tertentu. Jika diinginkan dapat juga menggunakan median dari cluster tersebut. sehingga rata-rata (mean) bukanlah satu-satunya ukuran yang dapat dipakai. (2) dimana : RR_{kk} = Rata-rata baru. N_k = Jumlah training pattern pada cluster (k). X_{nk} = Pola ke (n) yang menjadi bagian cluster (k).
5. kerjakan lagi pada objek dengan menggunakan pusat cluster. Jika pusat cluster tidak berubah maka proses clustering selesai. Atau, kembali ke langkah nomor 3 [10].

III. METODOLOGI PENELITIAN

Pada metodologi penelitian terdapat urutan kerangka kerja yang dilakukan. Kerangka kerja ini merupakan langkah-langkah yang dikerjakan pada penelitian. kerangka kerja yang digunakan pada penelitian ini adalah seperti terlihat pada gambar berikut.





Gambar 1. Kerangka Kerja Penelitian

Berikut penjelasan tahapan dalam penelitian :

1. Mengidentifikasi Masalah

Pada tahap ini untuk menentukan permasalahan yang terdapat pada penelitian, dengan cara mengumpulkan studi literature sebagai bahan melalui jurnal-jurnal terpublikasi nasional maupun internasional Untuk mengetahui permasalahan apa saja yang terjadi dan yang dimiliki oleh peneliti sehingga ditemukan pokok permasalahan yang akan diselesaikan kedepannya

2. Pengumpulan Data

Setelah mengidentifikasi permasalahan yang dimiliki, tahap selanjutnya adalah pengumpulan data. Pengumpulan data-data tersebut dari buku dan jurnal yang berhubungan TBC batam di BPS Riau. Dalam penelitian ini tahap pengumpulan data dilakukan beberapa cara yaitu :

a. Wawancara

Penelitian ini mendapatkan data-data dari BPS Riau dengan melalui wawancara langsung di kantor BPS Riau untuk mendapatkan informasi yang berkaitan terhadap Penderita TBC

b. Studi Pustaka

Pada penelitian juga menggunakan teknik pengumpulan data studi pustaka dilakukan dengan mempelajari teori-teori atau literatur dari buku-buku, jurnal, referensi-referensi yang berhubungan dengan K-Means untuk melengkapi data. Pengumpulan data mengenai penderita TBC

3. Analisis Data

Setelah pengumpulan data, tahap selanjutnya adalah Analisis Data. Pada tahap ini, akan dilakukan analisa terhadap apa-apa saja hal yang dibutuhkan dalam Data dan informasi yang dikumpulkan akan digunakan untuk mendukung penelitian dan menetapkan hasil apa yang akan diperoleh oleh analisa tersebut. Dalam tahap ini dibutuhkan kerja sama antara peneliti dengan BPS Riau sehingga hasil yang diinginkan sesuai dengan yang diharapkan.

4. Implementasi dengan K-means

Tahap selanjutnya adalah implementasi algoritma k-means untuk melakukan klasterisasi penderita TBC yang ada di wilayah Provinsi Riau.

5. Analisa Hasil

Setelah implementasi menggunakan K-means tersebut, diharapkan mampu memecahkan masalah tentang kesulitan dalam klasterisasi penderita TBC di Riau sehingga dapat membantu pemerintah dalam mengambil tindakan selanjutnya ke depan..

IV. HASIL DAN PEMBAHASAN

4.1 Data Selection

Pada tahap pemilihan data, data yang dipakai adalah data jumlah penderita TBC pada provinsi Riau. Data penderita TBC dapat dilihat pada Tabel. 1

Tabel 1 Data Penderita TBC Di Riau

No	Kota	Luas (Km2)	Jumlah Penderita

1	bengkalis	7774	11	4092
2	Dumai	2039	7	2923
3	Indragiri Hilir	11606	20	1469
4	Indragiri Hulu	7724	14	1369
5	Kampar Kuantan	11289	21	5052
6	Singingi	7656	15	2026
7	Meranti Pekanbaru	3714	9	1260
8	Pelalawan	632.3	12	8082
9	Rokan Hilir	12758	12	2783
10	Rokan Hilir	8852	15	3589
11	Hulu Rokan	7588	16	8617
12	Siak	8275	14	2790

4.2 Centroid Data

Pada implementasi algoritma K-means dapat menghasilkan nilai median atau centroid dari data. Proses pencarian nilai titik tengah dikerjakan dengan cara memilih nilai terbesar (maksimum) pada cluster tinggi (C1), nilai rata-rata (average) untuk cluster sedang (C2) dan nilai terkecil (minimum) untuk cluster rendah (C3). dan nilai centroid awal pada iterasi 1 diketahui pada tabel berikut:

Tabel 2 Hasil Perhitungan Jarak Data

	Centroid	
Max (C1)	12	12
Average (C2)	13.833	4.064
Min (C3)	6.500	3.606

4.3 Hasil Pelatihan

Tahapan berikut ini menggunakan algoritma K-Means untuk mengelompokkan data yang ada. Data yang ada akan dikelompokkan menjadi 3 kelompok. Adapun langkah dari pengelompokkan data adalah :

1. Menentukan titik pusat cluster secara acak,
2. Lakukan perhitung jarak setiap data yang ada terhadap setiap pusat cluster. Sehingga menghasilkan hasil perhitungan seperti berikut.

Tabel 3 Hasil Perhitungan Jarak Data

Kota	Pend erita	Keca matan	C1	C2	C3
bengkalis	4092	11	5,497	3,563	7,654
Dumai	2923	7	6,342	3,047	8,805
Indragiri Hilir	1469	20	7,506	2,958	10,240
Indragiri Hulu	1369	14	7,593	2,985	10,340
Kampar Kuantan Singingi	5052	21	4,897	4,183	6,712
Singingi	2026	15	7,047	2,908	9,690
Meranti Pekanbaru	1260	9	7,688	3,016	10,449
Pelalawan	8082	12	4,092	6,708	3,813
Rokan Hilir	2783	12	6,446	3,005	8,943
Rokan Hulu	3589	15	5,843	3,295	8,148
Siak	8617	16	4,162	7,192	3,324
	2790	14	6,440	3,004	8,935

Nilai data dapat menjadi bagian pada suatu cluster yang mempunyai jarak terkecil dari pusat clusternya. contohnya pada data pertama, jarak terkecil diperoleh pada cluster pertama, sehingga data pertama menjadi anggota dari cluster pertama. Demikian juga untuk data kedua, jarak terkecil ada pada cluster ketiga, maka data dapat masuk kepada cluster ketiga. hasil cluster setelah dilakukan perhitungan sebagai berikut.

Tabel 4 Hasil Perhitungan Jarak Data

Kota	Pender ita	Kecama tan	C 1	C 2	C 3
Bengkalis	4092	11	1		
Dumai	2923	7		1	
Indragiri Hilir	1469	20		1	
Indragiri Hulu	1369	14		1	
Kampar Kuantan Singingi	5052	21	1		
Meranti	1260	9		1	
Pekanbaru	8082	12			1
Pelalawan	2783	12		1	
Rokan Hilir	3589	15	1		
Rokan Hulu	8617	16			1
Siak	2790	14		1	

Pada tabel 4, diperoleh pengelompokan untuk iterasi 1: Cluster tingkat tinggi (C1) terdiri dari (C2) dan (C3)

Proses iterasi berjalan hingga nilai iterasi terakhir memiliki nilai yang mendekati dengan hasil iterasi sebelumnya. hasil nilai titik tengah atau centroid akan mengikuti dengan iterasi yang ada. Pada kegiatan ini proses akan dilanjutkan dengan memanfaatkan aplikasi RapidMiner.

4.4 Penerapan RapidMiner

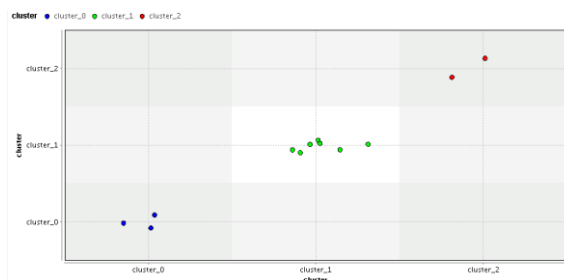
Hasil klastering pada data daerah penderita TB provinsi Riau dengan Rapid Miner :

1. Data view: ini data sheet di Result Perspective berfungsi untuk menampilkan data yang telah diolah berdasarkan dengan klasternya dari example set.

Row No.	id	cluster	Jumlah Pen.	Jumlah Kec.
1	1	cluster_0	4092	11
2	2	cluster_1	2923	7
3	3	cluster_1	1469	20
4	4	cluster_1	1369	14
5	5	cluster_0	5052	21
6	6	cluster_1	2026	15
7	7	cluster_1	1260	9
8	8	cluster_2	8082	12
9	9	cluster_1	2783	12
10	10	cluster_0	3589	15
11	11	cluster_2	8617	16
12	12	cluster_1	2790	14

Gambar 2. Data view

2. Flot view : merupakan sheet pada Result Perspective berfungsi sebagai menampilkan data yang selesai diolah secara keseluruhan lengkap berdasarkan klasternya dari contoh set (read excel) berupa bentuk diagram Scatter.



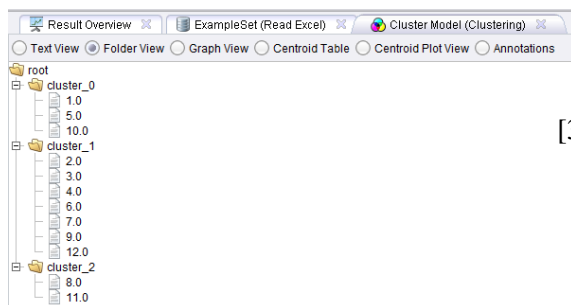
Gambar 3. Flot View

3. Text view : merupakan bagian untuk menampilkan basis data yang menampilkan hasil yang telah diolah secara keseluruhan lengkap dengan klasternya.

Cluster Model	
Cluster 0:	3 items
Cluster 1:	7 items
Cluster 2:	2 items
Total number of items: 12	

Gambar 4. Text View

4. Folder view : adalah bagian dari untuk menampilkan basis data yang telah diolah secara keseluruhan lengkap dengan clusternya.



Gambar 5. Folder view

V. KESIMPULAN

Penelitian ini telah selesai dilaksanakan, sehingga dapat ditarik beberapa kesimpulan yaitu, bahwa hasil dari pengujian yang dikerjakan pada penelitian ini, iterasi clustering pada data jumlah penderita TBC di Riau tersebut dapat diolah menggunakan aplikasi bantuan Rapid minner yang digunakan untuk menentukan dari nilai centroid pada 3 cluster yaitu cluster Rendah (C1), cluster sedang (C2) dan cluster Tinggi (C3). Dan dari 12 wilayah yang masuk kedalam cluster tingkat Rendah untuk penderita kasus TBC 3 wilayah , 7 wilayah cluster tingkat sedang dan 2 provinsi lainnya termasuk cluster tingkat Tinggi.

VI. SARAN

VII. DAFTAR PUSTAKA:

- [1] P. A. Apriadisiregar, F. P. Gurning, E. Eliska, and M. Y. Pratama, "Analysis of Factors Associated with Pulmonary Tuberculosis Incidence of Children in Sibuhuan General

Hospital," *J. Berk. Epidemiol.*, vol. 6, no. 3, p. 268, 2018, doi: 10.20473/jbe.v6i32018.268-275.

- [2] B. Maschinen, A. Investition, G. Beschaffungen, B. Ersatzbeschaffungen, and S. Mittelherkunft.

- [3] N. H. Muchtar, D. Herman, and Y. Yulistini, "Gambaran Faktor Risiko Timbulnya Tuberkulosis Paru pada Pasien yang Berkunjung ke Unit DOTS RSUP Dr. M. Djamil Padang Tahun 2015," *J. Kesehat. Andalas*, vol. 7, no. 1, p. 80, 2018, doi: 10.25077/jka.v7i1.783.

- [4] A. S. Chan, "Jurnal Ilmiah Informatika (JIF) Prediksi Kedatangan Wisatawan Pada Pariwisata Kota Batam Dengan Menggunakan Teknik Knowledge Data Discovery," pp. 1–6, 2018.

- [5] A. S. Chan, "Analytical Hierarchy Process Dan Fuzzy Topsis Pada Sistem Pendukung Keputusan," vol. 05, no. 01, pp. 1–14, 2018.

- [6] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.

- [7] T. Rismawan and D. S. Kusumadewi, "Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (Bmi) & Ukuran Kerangka," *Semin. Nas. Apl. Teknol. Inf.*, vol. 21, no. 01, pp. 1907–5022, 2008.

- [8] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-

- Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.
- [9] D. Alfiandi, Ernawati, and E. P. Purwandari, “Implementasi K-Means Clustering Dan Pemetaan Pemukiman Kumuh Di Kota Bengkulu Berbasis Web,” vol. 6, no. 2, 2018.
- [10] Y. R. Nasution *et al.*, “Penerapan Algoritma K-Means Clustering Pada Aplikasi Menentukan Berat Badan Ideal,” *J. Ilmu Komput. dan Inform.*, vol. 02, no. April, pp. 77–81, 2018.