

Implementasi Algoritma K-Means dan Fuzzy C-Means Clustering untuk Pengelompokan Penyebaran Penyakit Menular TBC Paru

A Comparison of K-Means and Fuzzy C-Means Clustering Algorithms for Clustering the Spread of Tuberculosis (TB) in the Lungs

¹Faradila Ramadani*, ²M. Afdal, ³Mustakim, ⁴Rice Novita
^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau

*e-mail: 11950320017@students.uin-suska.ac.id

(received: 15 June 2024, revised: 19 June 2024, accepted: 23 Jun 2024)

Abstrak

Tuberkulosis (TBC) merupakan penyakit menular dan menginfeksi individu dari segala rentang usia, termasuk bayi, anak-anak, remaja, dan lansia yang ditularkan melalui udara. Penyakit ini telah menyebar di berbagai wilayah Kabupaten Indragiri Hilir, sehingga penting untuk mengidentifikasi dan mengelompokkan daerah-daerah yang menjadi pusat penyebarannya. Tujuan penelitian ini untuk membantu pihak rumah sakit dalam melakukan penyuluhan di wilayah yang banyak terjangkit tuberkulosis. Penelitian ini memanfaatkan metode Data Mining dengan teknik klasterisasi algoritma K-Means dan Fuzzy C-Means berdasarkan data pasien Rumah Sakit Puri Husada Tembilahan dari tahun 2020 hingga 2023. Setelah beberapa percobaan, hasilnya dievaluasi menggunakan DBI, yang menunjukkan bahwa K-Means memberikan validitas terbaik dengan nilai 0,9146. Yang menunjukkan wilayah yang memiliki kasus tuberkulosis resiko tinggi penyakit TBC adalah tembilahan, dengan usia 55-64 dengan diagnose TBC komplikasi. Metode ini kemudian diimplementasikan dalam Sistem Informasi Clustering Tuberkulosis di RSUD Puri Husada Tembilahan dengan harapan dapat membantu pekerjaan di rumah sakit dalam mengurangi tingkat penyebaran penyakit di wilayah yang terjangkit tersebut.

Kata kunci: DBI, fuzzy c-means, klastering, k-means, tuberkulosis

Abstract

Tuberculosis (TB) is an airborne infectious disease that affects people of all ages, including infants, children, teenagers and the elderly. This disease is prevalent in different areas of Indragiri Hilir Regency, so it is important to identify and group the areas that are the focus of its spread. The purpose of this study is to help hospitals organize training in areas where tuberculosis is common. This study uses a data mining method with grouping techniques of K-Means and Fuzzy C-Means algorithms based on patient data from Puri Husada Tembilahan Hospital from 2020 to 2023. After several experiments, the results were evaluated with DBI, which showed that K- Means gave the best validity with a value of 0.9146. Which shows that the areas with high risk of TB are Tembilahans aged 55-64 who have been diagnosed with complicated TB. This method was then applied to the TB group information system of Puri Husada Tembilahan District Hospital in the hope that it could help the hospital reduce the spread of the disease in the affected area. Keywords: DBI, fuzzy c-means, clustering, k-means, tuberculosis.

Keywords: DBI, fuzzy c-means, clustering, k-means, tuberculosis

1. Pendahuluan

Tuberkulosis (TBC) merupakan penyakit menular dan menginfeksi individu dari segala rentang usia, termasuk bayi, anak-anak, remaja, dan lansia yang ditularkan melalui udara. Penyakit ini

<http://sistemasi.ftik.unisi.ac.id>

menyebabkan morbiditas dan mortalitas, menyebabkan kematian lebih dari 1 juta setiap tahunnya. TBC disebabkan oleh mikroorganisme bernama *Mycobacterium tuberculosis* [1]. Penularan TBC terjadi melalui udara dalam bentuk percikan dahak yang dilepaskan ketika seorang penderita batuk atau bersin [2]. Penyakit ini dapat menyebar secara cepat di individu imun rendah dan daya tahan tubuh yang lemah. Seseorang yang menderita tuberkulosis dapat menularkan infeksi ini kepada 1 dari 10 orang di sekitarnya [3].

Indonesia adalah salah 1 dari 5 negara yang mengalami tingkat kasus tuberkulosis paling tinggi di dunia, sebanyak 56% dari total kasus global, bersama dengan India, China, Filipina, dan Pakistan (WHO, 2019). Berdasarkan studi inventarisasi TB tahun 2017, angka kejadian TBC di Indonesia mencapai 318 kasus per 100.000 penduduk, yang setara dengan sekitar 842.000 kasus. Penelitian ini mengungkap adanya kecenderungan kurangnya pelaporan kasus tuberkulosis (TBC) di pusat kesehatan masyarakat sebesar 41% hingga 15%, sementara di fasilitas layanan kesehatan lainnya seperti rumah sakit, klinik, praktik dokter mandiri, dan laboratorium, tingkat under-reporting mencapai 71%. Prevalensi kasus TBC resisten obat multidrug-resistant (MDR) pada pasien baru adalah 1,4%, sedangkan pada pasien lama yang menjalani pengobatan ulang mencapai 13,1% [4].

Tuberkulosis juga merupakan penyakit menular mematikan, menempati urutan kedua sebagai penyebab kematian setelah virus COVID-19. Latar belakang riset ini adalah banyaknya masalah dan kasus tuberkulosis yang ditemukan di Provinsi Riau. Data yang dimanfaatkan dalam penelitian yaitu kasus TBC diperoleh dari Rumah Sakit Puri Husada di Indragiri Hilir, Provinsi Riau. Berdasarkan data yang di dapatkan pada 2020 terdapat 183 kasus tuberkulosis, dan pada tahun 2021 terdapat 127 kasus, dan pada tahun 2022 kasus TB meningkat kembali menjadi 153 dan pada tahun 2023 kasus bertambah menjadi 253.

Mengingat tingginya jumlah kasus tuberkulosis yang berbahaya bagi masyarakat Indonesia, terutama di Kabupaten Indragiri Hilir, diperlukan upaya pencegahan melalui analisis teknik data mining. Analisis ini akan dilakukan dengan memanfaatkan metode klusterisasi K-Means dan Fuzzy C-Means [5]. K-Means bisa menjadi pilihan yang efektif dalam menyelesaikan permasalahan. Data mining mencakup identifikasi struktur kumpulan data banyak dengan memanfaatkan algoritma seperti K-Means bisa digunakan untuk menganalisis data kesehatan masyarakat yang kompleks dan multidimensional [6]. Selain K-Means, Fuzzy C-means juga salah satu metode yang terkenal dalam analisis klaster. Ini adalah teknik klusterisasi yang umum digunakan dalam berbagai aplikasi klusterisasi. Dimana setiap data memiliki tingkat keanggotaan pada setiap klaster dengan pendekatan fuzzy [7].

Beberapa studi sebelumnya telah menyelidiki algoritma yang serupa. Sebagai contoh, dalam satu penelitian, K-Means dan Fuzzy C-Means diterapkan guna mengelompokkan data kasus Covid-19 di Kabupaten Indragiri Hilir. Hasilnya menunjukkan bahwa menggunakan nilai $K=3$, baik K-Means maupun Fuzzy C-Means menghasilkan klaster terbaik. Fuzzy C-Means terbukti sebagai metode terunggul dalam pengelompokan karena mendekati nilai validitas sebesar 1 [8]. Penelitian lain membandingkan algoritma K-Means dan Fuzzy C-Means dalam melakukan klusterisasi puskesmas menurut status gizi balita di Surabaya. Hasil penelitian menunjukkan bahwa nilai terbaik dari koefisien siluet untuk K-Means setelah proses normalisasi adalah 0,5183, sedangkan untuk Fuzzy C-Means setelah proses normalisasi adalah 0,4967. Hasil menunjukkan kinerja algoritma K-Means lebih optimal [9]. Penelitian lain menggunakan metode K-Means untuk klasifikasi penyakit ISPA dan berhasil mencapai tingkat akurasi yang tinggi serta efisiensi dalam mengidentifikasi gejala penyakit ISPA. Dalam pengujian digunakan validasi DBI, hasil menunjukkan bahwa klaster 1 memiliki nilai DBI -0,244, klaster 2 memiliki nilai DBI -0,250, dan klaster 3 memiliki nilai DBI -0,239. Karena nilai DBI klaster 3 lebih rendah, klaster ini dianggap optimal dalam konteks analisis tersebut [10].

Berdasarkan permasalahan yang telah diuraikan sebelumnya, peneliti memanfaatkan metode K-Means dan Fuzzy C-Means untuk melakukan klusterisasi penyakit TBC berdasarkan lokasi geografis. Tujuan penelitian ini untuk pengelompokan data ke dalam beberapa klaster yang sama untuk menerapkan, menguji, dan mengevaluasi kedua algoritma tersebut. Diharapkan hasil penelitian ini dapat membantu rumah sakit dan pemerintah dalam menangani penyebaran penyakit TBC.

2. Tinjauan Pustaka

Tinjauan pustaka berisi perbahaasan terkait penelitian yang menunjukkan poin-poin dari penelitian ini. Tinjauan pustaka diperoleh dari berbagai referensi seperti jurnal, buku, ataupun tesis, dan karya ilmiah lainnya.

2.1. Tuberculosis

Tuberkulosis (TBC) termasuk dalam 10 penyakit dengan tingkat fatalitas tertinggi di seluruh dunia. TBC disebabkan oleh infeksi bakteri *Mycobacterium tuberculosis*, yang bisa ditularkan lewat udara ketika individu yang terinfeksi batuk, bersin, atau berbicara, sehingga orang di sekitarnya dapat menghirup bakteri tersebut. Gejala TBC meliputi batuk yang berlangsung lebih dari tiga minggu, nyeri dada, batuk berdarah, kelemahan atau kelelahan yang persisten, penurunan berat badan, demam, serta keringat yang berlebihan di malam hari [11].

2.2. Data Mining

Teknik analisis data besar untuk mengidentifikasi pola-pola yang untuk mengetahui informasi yang tersembunyi dari dalam data tersebut merupakan pengertian dari Data mining. Data mining juga merupakan kombinasi dari metode statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengeksplorasi informasi, sehingga dapat digunakan untuk tujuan prediksi, pengambilan keputusan, dan optimisasi proses bisnis [12].

2.3. Clustering

Klaster mengacu pada pengelompokan entitas seperti catatan, pengamatan, atau objek yang membentuk kelas dengan kesamaan tertentu [13]. Klasterisasi merupakan tahap mengelompokkan entitas berdasarkan data yang ada guna mencerminkan hubungan antara objek tersebut. Prinsip utama dari klasterisasi adalah untuk meningkatkan keseragaman di antara individu yang sekelompok (klaster) dan mengurangi keseragaman antar kelompok [14].

2.4. K-Means

Algoritma K-Means adalah sebuah teknik klasterisasi non-hirarkis yang efisien dalam hal waktu dan mampu mengorganisir data dalam skala besar dengan kecepatan perhitungan yang relatif tinggi [15]. Prinsip utama dari teknik ini adalah menempatkan K pusat massa atau centroid dari sekumpulan data yang ada [16]. Dalam mengetahui jarak antara setiap data dengan setiap centroid dalam sebuah klaster, digunakan persamaan jarak Euclidean berikut ini.

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

2.5. Fuzzy C-Means

Jim Bezdek pada tahun 1981 mengenalkan metode Fuzzy C-Means yang merupakan teknik pengelompokan data di mana data dapat memiliki keanggotaan dalam beberapa klaster, yang ditentukan oleh nilai derajat keanggotaannya yang berkisar dari 0 hingga 1 [17]. FCM (Fuzzy C-Means) memiliki kemampuan mengetahui cluster tingkatan besar serta mengatakan ikatan antara model cluster yang berbeda. Peranan objektif ataupun peranan fitness bisa diminimalkan untuk memperoleh struktur klaster serta hasil klaster yang maksimal. [18]

2.6. Davies Boulden Index(DBI)

Indeks Davies-Bouldin (DBI) adalah sebuah teknik penilaian internal yang digunakan untuk mengevaluasi tingkat suatu teknik klasterisasi berdasarkan ukuran kohesi dan separasi. Kohesi merujuk pada tingkat keseragaman data dalam suatu klaster terhadap centroidnya, sedangkan separasi berkaitan dengan jarak antara centroid klaster yang berbeda [19]. DBI merupakan metode evaluasi yang menggunakan standar nilai indikator, di mana semakin rendah nilai DBI yang didapatkan (dalam

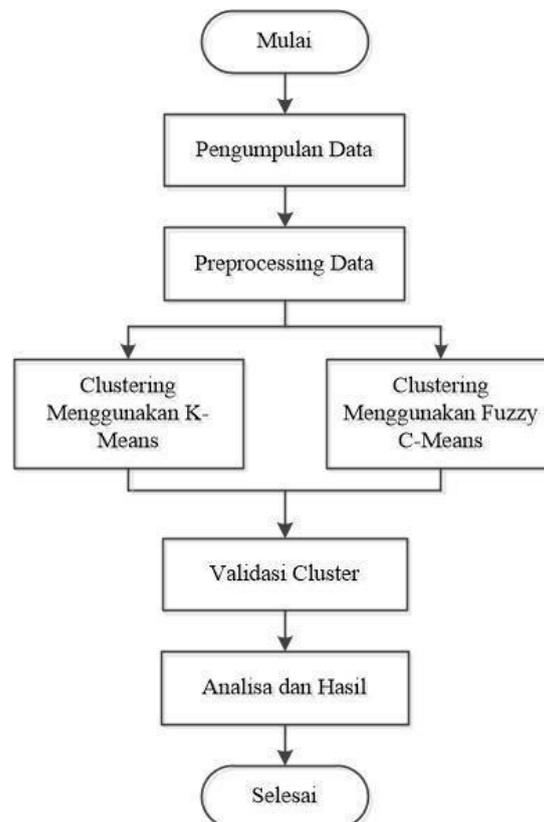
rentang non-negatif ≥ 0), semakin optimal kluster yang terbentuk dari proses pengelompokan yang dilakukan [20].

2.7. Penelitian Terkait

Penelitian ini dilakukan dengan merujuk dari beberapa penelitian terkait dengan penggunaan algoritma K-Means dan Fuzzy C-Means dalam melakukan clustering. Penelitian yang melakukan clustering dengan K-Means dilakukan oleh yanti puspita sari et al. (2024), dengan Judul *Implementasi Algoritma K-Means Untuk Clustering Penyebaran Tuberculosis Di Kabupaten Karawang* menghasilkan Penelitian dimulai dengan menentukan jumlah cluster optimal menggunakan metode elbow, yang mengindikasikan bahwa $K=3$ adalah pilihan terbaik. Kemudian, $K=3$ diuji dengan variasi random seed (S) 5, 10, dan 15. Hasil dari uji coba tersebut dievaluasi menggunakan SSE (Sum of Square Error). Clustering terbaik diperoleh dengan $K=3$ dan $S=10$, yang menghasilkan cluster 0 dengan 7 anggota, cluster 1 dengan 9 anggota, dan cluster 2 dengan 14 anggota. (2024)[21]. Dan penelitian yang dilakukan oleh la dodo et al (2023) dengan menggunakan algoritma Fuzzy C-Means dengan judul *Analysis Of Tuberculosis Disease Case Growth From Medical Record Data, Viewed Through Clustering Algorithms* menghasilkan alah satu cara untuk memahami pertumbuhan kasus tuberkulosis adalah melalui pengelompokan menggunakan Teknik Data Mining, khususnya beberapa algoritma clustering seperti k-means clustering, fuzzy c-means, dan Gaussian mixture. Teknik-teknik ini bertujuan untuk mengidentifikasi pertumbuhan kasus tuberkulosis berdasarkan rentang usia dan jenis kelamin. Oleh karena itu, hasil penelitian diharapkan dapat memberikan wawasan baru yang berharga bagi pembuat kebijakan dalam berbagai kapasitas, seperti langkah-langkah pencegahan, penyediaan fasilitas kesehatan, dan pertimbangan pengobatan.(2023)[22]

3. Metode Penelitian

Metode penelitian yaitu kuantitatif, yang diakui karena pendekatannya yang sistematis dan terorganisir secara baik. Pendekatan ini dikenal sebagai kuantitatif karena peneliti menggunakan data berupa nilai numerik dalam proses penelitiannya [23]. Dimulai dari tahap perencanaan hingga tahap dokumentasi yang tergambar dalam Gambar 1.



Gambar 1. Metodologi penelitian

3.1. Pengumpulan data

Data pasien penderita Tuberkulosis yang digunakan sebagai subjek analisis. Data tersebut diperoleh dari Rumah Sakit Puri Husada yang terletak di Jalan Veteran No. 52, Tembilahan, Indragiri Hilir, Riau. Data yang digunakan mencakup informasi mengenai pasien yang mengidap Tuberkulosis paru dari tahun 2020 hingga 2023.

3.2. Tahap Preprocessing

Pada penelitian ini dilakukan beberapa tahapan dalam memproses data, yaitu dengan melakukan [24]:

1. **Data selection** merupakan tahap untuk memilih data cocok dari basis data sesuai dengan tujuan analisis.
2. **Cleaning data** adalah proses untuk menghilangkan duplikat data dan konsistensi data yang tidak sesuai.
3. **Transformasi data** merupakan langkah untuk mengubah format data agar sesuai dengan proses mining yang dilakukan.

3.3. Proses Clustering

Proses pengelompokan menggunakan algoritma K-means dan algoritma Fuzzy C Means (FCM). Permodelan algoritma dilakukan dengan percobaan 2 cluster sampai dengan 5 cluster. Tujuan dari dilakukannya analisis clustering adalah agar data yang diasosiasi menjadi lebih kecil, sehingga pola yang dihasilkan nantinya dapat lebih akurat.

3.4. Validitas Cluster

Proses Validitas Cluster adalah membandingkan algoritma K-means dan FCM dengan dilakukannya uji validasi cluster untuk mencari nilai validitas cluster terbaik dari kedua algoritma yang telah dilakukannya proses clustering. Uji validitas cluster menggunakan metode Davies Boulden Index (DBI). Hasil nilai validitas cluster tertinggi merupakan cluster yang memiliki kualitas cluster terbaik.

4. Hasil dan Pembahasan

Bagian ini mencakup semua tahapan analisis penelitian, mulai dari pengumpulan data, preprocessing data, pengelompokan (clustering), hingga hasil analisis dari penelitian tersebut.

4.1. Pengumpulan Dan Preprocessing data

Data penelitian ini diperoleh dari Rumah Sakit Puri Husada Tembilahan selama periode tahun 2020 hingga 2023, dengan jumlah total pasien yang terkonfirmasi menderita TBC mencapai 668 orang yang ditampilkan pada Tabel 1. Data yang telah melalui proses pembersihan data, yang ditampilkan pada Tabel 2

Tabel 1. Data Tuberculosis

No	No MR	Alamat	Usia	Diagnosa
1	276854	Batang Tuaka	54	TBC Paru
2	381856	Enok	49	TB MDR Lesi Luas + Anemia Berat + Dyspepsia
3	319590	Gaung	33	TB Paru Putus Obat + Asma
4	120479	gaung anak serka	42	TB Paru Kiri Luas + Cachexia + Dyspepsia
5	358598	Kempas	31	Lung Tuberculosis + Dispnoe
6	389438	Kemuning	23	Primary Lung Tuberculosis + Malnutrisi + B20
7	152262	Keritang	33	TB MDR
9	322113	kuala indragiri	65	TB Paru Luas + Pneumonia + Malnutrisi
10	152265	Mandah	32	TB MDR

688	333388	Jambi	33	lung TB + hemaptoe

Tabel 2. Hasil data cleaning

No	No MR	Alamat	Usia	Diagnosa
1	276854	Batang Tuaka	54	TBC Paru
2	381856	Enok	49	TB MDR Lesi Luas + Anemia Berat + Dyspepsia
3	319590	Gaung	33	TB Paru Putus Obat + Asma
4	120479	gaung anak serka	42	TB Paru Kiri Luas + Cachexia + Dyspepsia
5	358598	Kempas	31	Lung Tuberculosis + Dispnoe
6	389438	Kemuning	23	Primary Lung Tuberculosis + Malnutrisi + B20
7	152262	Keritang	33	TB MDR
9	322113	kuala indragiri	65	TB Paru Luas + Pneumonia + Malnutrisi
10	152265	Mandah	32	TB MDR
...
641	339474	Tempuling	23	Primary Respiratory Tuberculosis + Pneumonia Efusi Pleura

4.2. Normalisasi Data

Normalisasi yang digunakan dalam analisis data dijadikan objek penelitian menggunakan metode normalisasi min-max. Hasil normalisasi data tersebut terdokumentasi dalam Tabel 3.

Tabel 3. Data pasien TBC hasil normalisasi data

No	No MR	Alamat	Usia	Diagnosa
1	276854	0,894737	0,333333	0,045455
2	381856	0,789474	0	0,863636
3	319590	0,894737	0,833333	0,863636
4	120479	0	1	0,863636
5	358598	0,578947	0,666667	0,090909
6	389438	0,894737	1	0,045455
7	152262	0,421053	0,5	0,863636
9	322113	0,894737	0,166667	0,045455
10	152265	0,526316	0,666667	0,863636
...
641	339474	0,789474	0,666667	0,863636

4.3. Klasterisasi dengan algoritma

Algoritma K-Means dan Fuzzy C-Means yang digunakan dalam melakukan klusterisasi pada RStudio dengan data yang telah dinormalisasi pada tahap sebelumnya. Percobaan klusterisasi dilakukan untuk menentukan jumlah kluster optimal menggunakan algoritma K-Means dan Fuzzy C-Means.

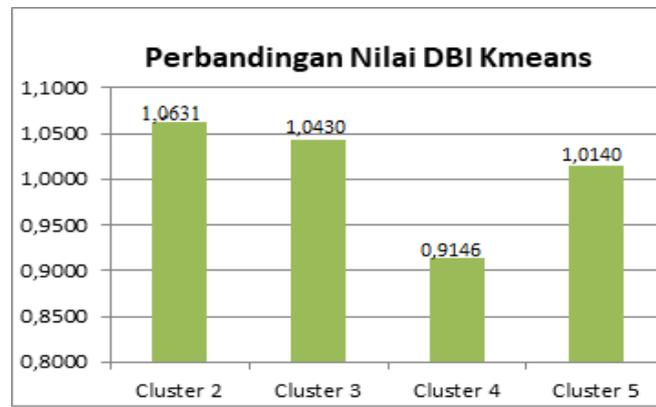
4.3.1. K-Means

Pada eksperimen ini, dilakukan pengujian dengan lima kluster bertujuan untuk mendapatkan cluster optimal. Tabel 4 menampilkan hasil percobaan dengan 5 cluster.

Tabel 4. Hasil clusterisasi K-Means

	Kriteria	CLUSTER			
		C2	C3	C4	C5
Kecamatan	Batang tuaka	0	0	0	16
	Concong	0	0	0	7
	Enok	0	0	0	14
	Gaung	0	0	0	14
	Gaung Anak Serka	0	0	0	6
	Kateman	0	0	0	2
	Kempas	0	1	0	13
	Kemuning	0	0	0	3
	Keritang	0	7	0	15
	Kuala Indragiri	4	4	0	3
	Mandah	11	3	0	0
	Pelangiran	8	21	3	0
	Pulau Burung	1	1	11	0
	Reteh	5	4	14	0
	Sungai Batang	0	1	5	0
	Tanah Merah	8	3	17	0
	Tanah Merah	2	2	2	0
	Teluk Belengkong	30	27	62	0
	Tembilahan	12	14	27	0
	Tembilahan Hulu	7	3	20	0
usia	Tempuling	0	1	2	0
	0-14	0	1	2	0
	15-24	0	22	17	3
	25-34	0	43	20	8
	35-44	0	25	22	14
	45-54	30	0	33	26
	55-64	29	0	41	17
≥65	29	0	25	25	
Diagnosa	Hemoptis Ec TB Paru	16	26	0	15
	Lung Tuberculosis (TBC)	65	44	0	62
	Primary Respiratory TB	2	6	0	3
	Sequele Of TB	0	0	0	3
	Susp TB Paru	1	5	0	3
	TB Konfirmasi Bakteriologi	0	1	0	0
	TB MDR	3	7	0	6
	TB Millier	1	0	0	1
	TB Paru dalam Pengobatan	0	1	1	0
	TB Paru Kiri Luas	0	1	1	0
	TB Paru Lama Luas	0	0	1	0
	TB Paru Lesi Luas	0	0	9	0
	TB Paru Relaps	0	0	1	0
	TB Dengan Komplikasi	0	0	145	0
	TB MDR+Hemaptoe	0	0	1	0
TB Paru Relaps+Hemaptoe	0	0	1	0	

Setelah proses pengelompokan dilakukan pada setiap kriteria, dilakukan evaluasi validitas kluster untuk menemukan kluster yang optimal. Evaluasi validitas kluster menggunakan Davies-Bouldin Index untuk algoritma K-Means. Hasil evaluasi validitas kluster tersebut ditampilkan pada Gambar 2.



Gambar 1. Perbandingan DBI

Gambar 2 menampilkan Nilai DBI terbaik pada cluster 4 dengan nilai 0.9146 yang terletak pada nilai K= 4, maka dapat diketahui bahwa kecamatan Indragiri Hilir yang memiliki resiko tinggi penyakit TBC adalah tembilahan, dengan usia 55-64 dan dengan diagnosa TBC dengan komplikasi.

4.3.2. Fuzzy C-Means

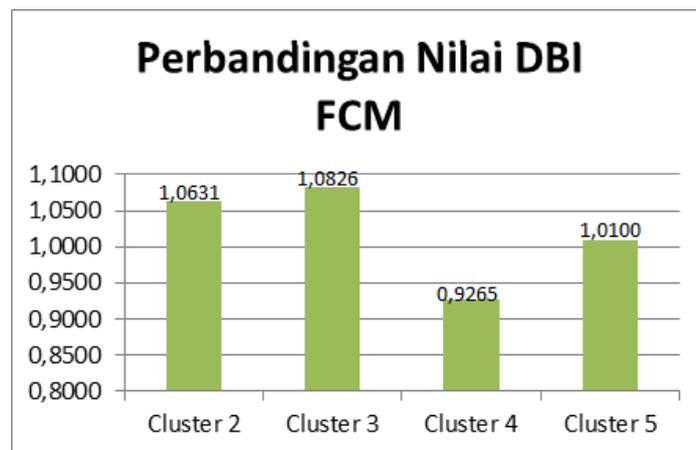
Dalam eksperimen ini, dilakukan klusterisasi menggunakan 5 kluster dengan nilai pangkat (w) sebesar 2 dan jumlah iterasi maksimum (Maxiter) sebanyak 1000. Detail eksperimen ini terdokumentasi dalam Tabel 5.

Tabel 5. Hasil klusterisasi FCM

Kriteria	CLUSTER			
	C2	C3	C4	C5
Batang tuaka	0	0	16	0
Concong	0	0	7	0
Enok	0	0	14	0
Gaung	0	0	14	0
Gaung Anak Serka	0	0	6	0
Kateman	0	0	2	0
Kempas	3	0	11	0
Kemuning	0	0	3	0
Keritang	7	0	15	0
Kecamatan Kuala Indragiri	4	0	7	0
Mandah	3	0	2	9
Pelangiran	21	19	0	8
Pulau Burung	1	11	0	5
Reteh	4	14	0	5
Sungai Batang	1	4	0	0
Tanah Merah	3	17	0	8
Teluk Belengkong	2	2	0	2
Tembilahan	27	51	0	30
Tembilahan Hulu	14	27	0	12
Tempuling	3	20	0	7
usia 0-14	1	2	0	0
15-24	22	21	3	0
25-34	45	25	6	0
35-44	25	27	14	0
45-54	36	0	28	28
55-64	0	41	17	29

Kriteria	CLUSTER			
	C2	C3	C4	C5
≥65	0	25	29	25
Hemoptis Ec TB Paru	26	0	17	14
Lung Tuberculosis (TBC)	46	0	64	61
Primary Respiratory TB	6	0	3	2
Sequele Of TB	0	0	3	0
Susp TB Paru	5	0	3	1
TB Konfirmasi Bakteriologi	1	0	0	0
TB MDR	7	0	6	3
TB Millier	0	0	1	1
Diagnosa TB Paru dalam Pengobatan	1	1	0	0
TB Paru Kiri Luas	1	1	0	0
TB Paru Lama Luas	0	1	0	0
TB Paru Lesi Luas	0	10	0	0
TB Paru Relaps	0	1	0	0
TB Dengan Komplikasi	0	160	0	0
TBC Kelenjar	0	1	0	0
TB MDR+Hemoptoe	0	1	0	0
TB Paru Relaps+Hemoptoe	0	1	0	0

Setelah tahap pengelompokan dilakukan pada setiap kriteria, dilakukanlah validitas klaster pada setiap kriteria pengelompokan dievaluasi untuk mencari klaster optimal menggunakan Davies-Bouldin Index untuk algoritma Fuzzy C-Means. Hasil evaluasi validitas klaster ini ditampilkan pada Gambar 3.

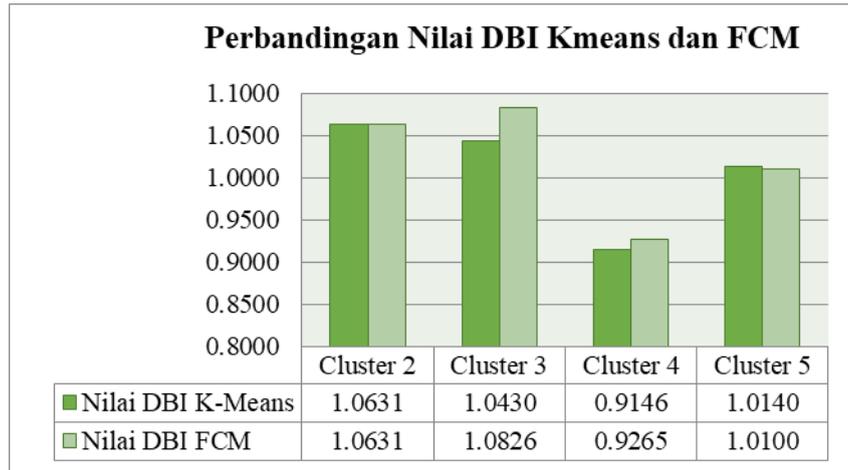


Gambar 3. Perbandingan FCM DBI

Dapat dilihat dari gambar 3 Nilai DBI terbaik pada 0,9265 yang terletak pada K=4, maka dapat diketahui bahwa kecamatan Indragiri hilir yang memiliki resiko tinggi penyakit TBC adalah maka dapat diketahui bahwa kecamatan indragirihilir yang memiliki resiko tinggi penyakit TBC adalah batang tuaka, dengan usia 65 tahun keatas dan dengan diagnosa TBC.

4.4. Perbandingan Validitas Cluster

Setelah melakukan proses klasterisasi menggunakan 2 algoritma, yaitu K-Means dan Fuzzy C-Means, dilakukan percobaan klaster dari 2 hingga 5, Langkah selanjutnya yaitu memilih klaster optimal untuk menguji validitas antara K-Means dan FCM pada setiap eksperimen. Detail perbandingan nilai validitas algoritma dapat ditemukan dalam Gambar 4.



Gambar 2. Perbandingan nilai DBI K-Means dn FCM

Pada Gambar 4, K-Means memperoleh nilai validitas tertinggi sebesar 0,9146, sementara Fuzzy C-Means mencapai 0,9265. Menurut kriteria validasi, semakin mendekati nilai 0 menunjukkan kualitas kluster yang lebih baik. Dengan demikian, K-Means mendapatkan validitas kluster yang lebih rendah daripada Fuzzy C-Means dengan nilai 0,9146.

4.5. Implementasi Sistem

Selanjutnya implementasi yaitu tahapan perancangan dan pembuatan aplikasi Sistem Informasi Clustering Tuberkulosis pada RSUD Puri Husada Tembilahan dengan mengaplikasikan algoritma K-Means pada sistem dengan harapan dapat mempermudah pekerjaan pada RSUD Puri Husada Tembilahan dalam melakukan pemeriksaan tuberkulosis pada pasien. Algoritma K-Means di pilih karena berdasarkan hasil perhitungan validitas K-Means memiliki nilai validitas cluster paling baik dibandingkan Fuzzy C-means, sehingga pada system informasi clustering tuberculosis ini hanya menggunakan algoritma K-Means.

a. Tampilan menu centeroid

Gambar 9 menunjukkan tampilan dari menu centeroid yang berisi tabel hasil clustering penderita tuberkulosis pada RSUD Puri Husada Tembilahan. Halaman tersebut menampilkan Diagnosa dari setiap masing-masing cluster dan juga menampilkan hasil dari cluster tersebut.

Hasil Clustering	No.MR	Nama Kecamatan	Usia (Thn)	Diagnosa	C1	C2	C3	C4	C5	Jar Ter
Cluster 1	329729	Batang Tuaka	35-44	Lung Tuberculosis (TBC)	5.15923799	18.92287430	17.95302008	24.13335453	14.54588267	5.1
Cluster 1	271743	Batang Tuaka	>=65	Lung Tuberculosis (TBC)	5.34021878	19.04014631	18.06728895	24.24300725	14.79148750	5.3
Cluster 1	331732	Enok	15-24	TB MDR	6.37579302	13.80680164	17.05283349	19.39996909	13.49915933	6.3
Cluster 1	315865	Gaung	55-64	Lung Tuberculosis (TBC)	2.19880346	18.17010103	14.98244740	22.21303223	11.67005152	2.1
Cluster 1	327380	Enok	25-34	Lung Tuberculosis (TBC)	3.73512204	18.42093297	16.03802139	22.85373056	12.61150675	3.7
Cluster 1	316223	Kempas	45-54	Sequele Of TB	1.91956679	15.82688762	12.16970542	18.78956627	8.68468206	1.9
Cluster 1	316613	Keritang	55-64	Hemoptis Ec TB Paru	3.38897280	18.93793472	10.01903838	20.45869010	6.92439907	3.3
Cluster 1	316753	Keritang	35-44	Lung	3.23804519	17.91740974	9.97134545	19.51892417	6.56197398	3.2

Gambar 3. Tampilan menu clustering

b. Visualisasi usia

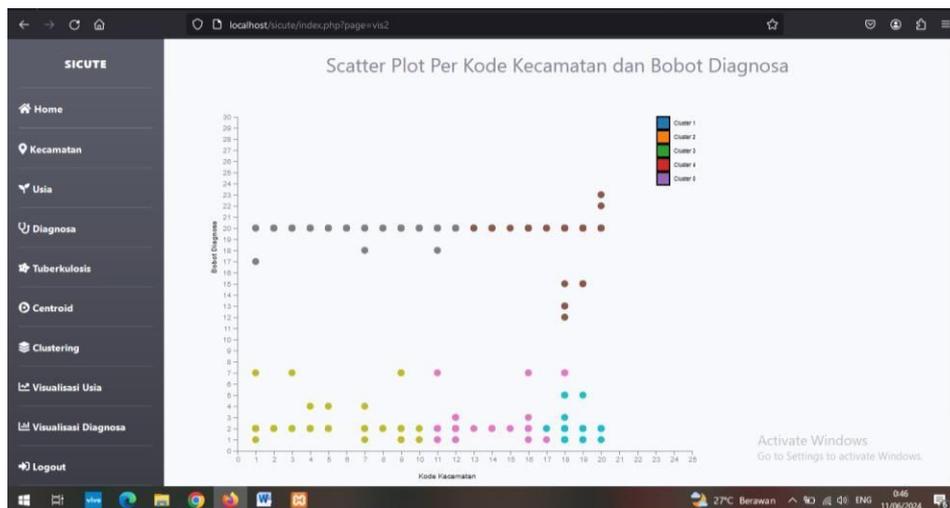
Berikut ini adalah visualisasi usia berdasarkan clusternya yang dapat dilihat pada Gambar 10. Dimana halaman tersebut menampilkan hasil Scatter Plot yang berisi Kode Kecamatan dan Bobot Usia berdasarkan hasil cluster penyakit tuberculosis di RSUD Puri Husada Tembilahan.



Gambar 4. Visualisasi usia

c. Visualisasi diagnosa

Menu visualisasi diagnose berisikan visualisasi dari hasil cluster berupa kecamatan dan diagnose dari hasil pemeriksaan. Hasil Visualisasi diagnose dapat dilihat pada Gambar 11.



Gambar 5. Visualisasi Diagnosa

5. Kesimpulan

Berdasarkan temuan dari penelitian ini, teknik K-Means dan Fuzzy C-Means telah berhasil diterapkan guna mengklasifikasikan pola kasus tuberculosis di Kabupaten Indragiri Hilir. Dari analisis yang dilakukan, diketahui bahwa algoritma terbaik yaitu K-Means menghasilkan nilai DBI sebesar 0,9146, sedangkan Fuzzy C-Means hanya memperoleh nilai DBI sebesar 0,9265. Hasil klusterisasi menghasilkan kecamatan Tembilahan dan Tembilahan Hulu memiliki risiko penyebaran tuberculosis yang tinggi, terutama pada kelompok usia 55-64 tahun dengan diagnosis Tuberculosis komplikasi. Dari hasil clustering diperoleh algoritma terbaik K-Mean yang kemudian dibangun aplikasi Sistem Informasi Clustering Tuberkolosis pada RSUD Puri Husada Tembilahan dengan mengaplikasikan algoritma K-Means pada sistem. Sehingga diharapkan aplikasi tersebut nantinya dapat digunakan di rumah RSUD Puri Husada Tembilahan untuk mempermudah proses clustering penyakit Tuberculosis (TBC).

Referensi

- [1] B. Yanti, "Penyuluhan Pencegahan Penyakit Tuberkulosis (TBC) Era New Normal," *Martabe J. Pengabd. Kpd. Masy.*, vol. 4, no. 1, p. 325, Mar. 2021, doi: 10.31604/jpm.v4i1.325-332.
- [2] D. Andriani and S. Sukardin, "Pengetahuan dan Sikap Keluarga Dengan Pencegahan Penularan Penyakit Tuberculosis (TBC) Di Wilayah Kerja Puskesmas Penana'e Kota Bima," *J. Ilm. Ilmu Keperawatan Indones.*, vol. 10, no. 03, pp. 72–80, Sep. 2020, doi: 10.33221/jiiki.v10i03.589.
- [3] A. Sejati and L. Sofiana, "Faktor-Faktor Terjadinya Tuberkulosis," *J. Kesehat. Masy.*, vol. 10, no. 2, p. 122, Jan. 2015, doi: 10.15294/kemas.v10i2.3372.
- [4] A. H. Mahpudin and R. Mahkota, "Faktor Lingkungan Fisik Rumah, Respon Biologis dan Kejadian TBC Paru di Indonesia," *Kesmas Natl. Public Heal. J.*, vol. 1, no. 4, p. 14, Feb. 2007, doi: 10.21109/kesmas.v1i4.297.
- [5] N. Nurfadilla, M. Afdal, I. Permana, and Z. Zarnelly, "Comparison of Data Mining Algorithm For Clustering Patient Data Human Infectious Diseases," *J. Tek. Inform.*, vol. 4, no. 5, pp. 1127–1134, Oct. 2023, doi: 10.52436/1.jutif.2023.4.5.983.
- [6] T. R. A. Ridho, M. Fakhriza, and A. Ikhwan, "Penerapan Algoritma k-Means Dalam Mengelompokkan Penyebaran Penyakit Di Kecamatan Sei Balai," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 8, no. 1, pp. 1–7, Jan. 2024, doi: 10.59697/jtik.v8i1.487.
- [7] N. D. Aina Latifa Riyana Putri, "Analisa perbandingan k-means dan fuzzy c-means dalam pengelompokan daerah penyebaran COVID-19 Indonesia," *Anal. perbandingan k-means dan fuzzy c-means dalam pengelompokan Drh. penyebaran COVID-19 Indones.*, vol. Vol 10 No, 2021, doi: <https://doi.org/10.15294/ujm.v10i2.50433>.
- [8] S. F. Octavia and M. Mustakim, "Penerapan K-Means dan Fuzzy C-Means untuk Pengelompokan Data Kasus Covid-19 di Kabupaten Indragiri Hilir," *Build. Informatics, Technol. Sci.*, vol. 3, no. 2, pp. 88–94, Sep. 2021, doi: 10.47065/bits.v3i2.1005.
- [9] A. Kurnia, "Perbandingan Algoritma K-Means dan Fuzzy C-Means Untuk Clustering Puskesmas Berdasarkan Gizi Balita Surabaya," *J. Process.*, vol. 18, no. 1, Apr. 2023, doi: 10.33998/processor.2023.18.1.696.
- [10] I. Romli, "Penerapan Data Mining Menggunakan Algoritma k-Means untuk Klasifikasi Penyakit Ispa," *Indones. J. Bus. Intell.*, vol. 4, no. 1, p. 10, Jun. 2021, doi: 10.21927/ijubi.v4i1.1727.
- [11] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [12] R. Azhari, D. Hartama, M. R. Lubis, D. F. Nasution, and A. P. Windarto, "Analisis Penerapan

- Data Mining Terhadap Kasus Positif Covid-19 Menggunakan Metode K-Means Clustering,” *J. Informatics, Electr. Electron. Eng.*, vol. 3, no. 2, pp. 221–235, Dec. 2023, doi: 10.47065/jieeee.v3i2.1760.
- [13] D. Ariyanto, “Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut,” *J. Sistim Inf. dan Teknol.*, pp. 13–18, Feb. 2022, doi: 10.37034/jsisfotek.v4i1.117.
- [14] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, p. 25, Aug. 2021, doi: 10.33365/jtk.v15i2.1162.
- [15] Mustakim, U. R. Gurning, A. P. Pristiawati, A. Dina, Putri, and Nabillah, “Pengelompokan Data Loyalitas Pelanggan Model RFM pada Produk Ms Glow Dan Scarlett dengan Algoritma Fuzzy C-Means,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind. Fak. Sains dan Teknol. UIN Sultan Syarif Kasim Riau*, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/19205/8332>
- [16] A. Pramudya, I. Maulana, and R. Mayasari, “Pengelompokan Hasil Belajar Siswa SDN Tunas Jaya Dengan Algoritma k-Means,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3960–3967, Jan. 2024, doi: 10.36040/jati.v7i6.7970.
- [17] B. A. Yudano, U. Indahyanti, and U. K. Nisak, “Pemanfaatan Data Rekam Medis dalam Menentukan Cluster Penyakit Melalui Data Mining di RS Aisyiyah Siti Fatimah Tulangan Sidoarjo,” *Care J.*, vol. 2, no. 1, pp. 14–23, Dec. 2022, doi: 10.35584/carejournal.v2i1.98.
- [18] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, “Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang,” *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 229, Nov. 2020, doi: 10.35314/isi.v5i2.1457.
- [19] L. Dodo, N. S. Fatonah, G. Firmansyah, and H. Akbar, “*Analysis of Tuberculosis Disease Case Growth From Medical Record Data, Viewed Through Clustering Algorithms (Case Study: Islamic Hospital Bogor)*,” *J. Indones. Sos. Sains*, vol. 4, no. 09, pp. 915–927, Sep. 2023, doi: 10.59141/jiss.v4i09.884.
- [20] H. Sulastri and A. I. Gufroni, “Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, Sep. 2017, doi: 10.25077/TEKNOSI.v3i2.2017.299-305.