

Analisis Pasien Tuberkulosis Menggunakan Metode *Fuzzy C-Means* Berbasis *Python*

Muhammad Fakhrol Rozi¹, Reny Wahyuning Astuti², Sukma Puspitorini³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Nurdin Hamzah,
Jambi, Indonesia

e-mail: ¹69mfrozi@gmail.com, ²r3ny4stuti@gmail.com, ³pijetsukm4@gmail.com

Abstrak

Puskesmas Simpang Kawat merupakan fasilitas pelayanan kesehatan yang salah satunya berperan dalam penanggulangan penyakit Tuberkulosis. Untuk membantu puskesmas dalam mengambil keputusan secara praktis dengan menggunakan keputusan berbasis data, diperlukan klasterisasi data pasien yang dapat membantu mengidentifikasi kelompok pasien dengan karakteristik tertentu. Penelitian ini menggunakan metode Fuzzy C-Means (FCM) yang diimplementasikan ke dalam pemrograman Python. Data pasien yang digunakan mencakup variabel umur, jenis kelamin, klasifikasi berdasarkan riwayat pengobatan sebelumnya, terduga Tuberkulosis Resisten Obat, hasil pemeriksaan foto toraks, hasil pemeriksaan tuberkulin, riwayat HIV dan riwayat diabetes. Metode FCM digunakan karena kemampuannya dalam mengelompokkan data dengan memberikan nilai derajat keanggotaan pada setiap cluster, sehingga memungkinkan analisis yang lebih fleksibel. Hasil penelitian menunjukkan bahwa data pasien dapat dikelompokkan ke dalam tiga cluster, yaitu: cluster rendah, cluster sedang dan cluster tinggi. Dengan demikian, metode FCM terbukti mampu dalam mengelompokkan pasien Tuberkulosis, sehingga dapat menjadi dasar untuk perencanaan intervensi medis, prioritas pemantauan, serta alokasi sumber daya di Puskesmas Simpang Kawat.

Kata kunci: *Fuzzy C-Means, Klasterisasi, Pasien, Python, Tuberkulosis*

Abstract

The Simpang Kawat Community Health Center is a healthcare facility that plays a role in the prevention and control of tuberculosis. To assist the health center in making practical decisions based on data, it is necessary to cluster patient data to help identify groups of patients with specific characteristics. This study uses the Fuzzy C-Means (FCM) method implemented in Python programming. The patient data used includes variables such as age, gender, classification based on previous treatment history, suspected drug-resistant tuberculosis, chest X-ray results, tuberculin test results, HIV history, and diabetes history. The FCM method was used because of its ability to group data by assigning membership values to each cluster, allowing for more flexible analysis. The results of the study show that patient data can be grouped into three clusters, namely: low cluster, medium cluster, and high cluster. Thus, the FCM method has been proven to be capable of grouping tuberculosis patients, thereby providing a basis for medical intervention planning, monitoring priorities, and resource allocation at the Simpang Kawat Community Health Center.

Keywords: *Clustering, Fuzzy C-Means, Patient, Python, Tuberculosis*

1. Pendahuluan

Indonesia menempati peringkat ketiga dengan jumlah kasus Tuberkulosis (TB) tertinggi di dunia setelah India dan Cina, dengan estimasi penderita mencapai 845.000 jiwa pada tahun 2019, meningkat dari 842.000 jiwa pada tahun sebelumnya [1]. TB merupakan penyakit menular yang disebabkan oleh *Mycobacterium Tuberculosis* dan dapat menyerang semua

kelompok usia melalui penularan udara saat penderita batuk atau bersin [2][3]. Kasus TB lebih tinggi pada laki-laki dibandingkan perempuan, yang dipengaruhi oleh faktor risiko seperti lingkungan tidak sehat dan kebiasaan merokok [4]. Meski berbahaya, penyakit ini sering dianggap sepele sehingga banyak penderita terlambat mencari pengobatan.

Di Puskesmas Simpang Kawat, telah ada program pencegahan dan pengendalian TB, namun pengelolaan data pasien masih bersifat manual melalui pencatatan sederhana. Hal ini menyulitkan tenaga kesehatan dalam melakukan pengelompokan pasien berdasarkan tingkat keparahan atau kategori tertentu, sehingga strategi penanganan dan pencegahan belum optimal. Diperlukan sistem analisis yang mampu mengelompokkan pasien dengan karakteristik serupa agar pelayanan lebih efektif dan efisien.

Sejumlah penelitian sebelumnya telah menerapkan metode klusterisasi untuk analisis data kesehatan, namun sebagian besar masih didominasi oleh penggunaan *K-Means*. Penelitian oleh Setyaningtyas et al. (2022) menunjukkan bahwa *K-Means* memiliki kelemahan karena bersifat *hard clustering*, sehingga setiap objek hanya dapat masuk ke dalam satu *cluster* secara tegas [5]. Berbeda dengan *K-Means*, metode *Fuzzy C-Means* memberikan fleksibilitas melalui pemberian derajat keanggotaan pada setiap cluster, sehingga lebih sesuai digunakan untuk data medis yang umumnya memiliki karakteristik dan kondisi pasien yang bervariasi [6].

Beberapa studi telah memanfaatkan *Fuzzy C-Means* dalam bidang kesehatan, di antaranya Pradana et al. (2021) yang berhasil menerapkan *Fuzzy C-Means* dalam prediksi kebutuhan pupuk tanaman padi berbasis citra dengan hasil segmentasi yang akurat [7]. Namun, penelitian terkait penggunaan *Fuzzy C-Means* untuk klusterisasi data medis, khususnya penyakit Tuberkulosis, masih sangat terbatas dan belum ditemukan studi yang secara spesifik menerapkan *Fuzzy C-Means* untuk pengelompokan karakteristik pasien Tuberkulosis di fasilitas layanan kesehatan tingkat pertama seperti puskesmas. Oleh karena itu, penelitian ini dilakukan untuk mengisi gap tersebut dengan mengimplementasikan metode *Fuzzy C-Means* berbasis *Python* pada data pasien Tuberkulosis, sehingga diharapkan dapat memberikan kontribusi pada pengembangan analisis data kesehatan berbasis *data mining*.

Meskipun berbagai penelitian telah memanfaatkan metode *Fuzzy C-Means* untuk proses klusterisasi pada berbagai bidang, penerapannya pada kasus kesehatan, khususnya Tuberkulosis di tingkat layanan kesehatan dasar, masih sangat terbatas. Sebagian besar penelitian yang menerapkan *Fuzzy C-Means* dalam bidang kesehatan lebih berfokus pada segmentasi citra medis, seperti citra X-ray atau CT-Scan [8], serta analisis *cluster* pasien dalam skala data besar di rumah sakit atau level nasional [9]. Namun, belum banyak penelitian yang mengimplementasikan *Fuzzy C-Means* pada data Tuberkulosis dengan skala data kecil hingga menengah di fasilitas kesehatan tingkat pertama seperti puskesmas, yang karakteristik datanya cenderung heterogen dan terbatas.

Selain itu, penelitian terkait Tuberkulosis pada puskesmas umumnya masih menggunakan pendekatan statistik deskriptif atau metode klasifikasi seperti *Naïve Bayes* atau *K-Means* [10], dan jarang memanfaatkan metode *Fuzzy C-Means* yang mampu memberikan representasi lebih realistis terhadap karakteristik pasien yang tidak sepenuhnya homogen. Dengan demikian, terdapat kesenjangan penelitian pada aspek:

1. Kurangnya penerapan *Fuzzy C-Means* dalam analisis kluster pasien Tuberkulosis di puskesmas,
2. Minimnya studi yang membandingkan efektivitas *Fuzzy C-Means* untuk pengelompokan risiko Tuberkulosis pada data klinis berskala kecil-menengah,
3. Perlunya pendekatan berbasis data mining untuk mendukung pengambilan keputusan klinis di layanan primer.

Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan menerapkan metode *Fuzzy C-Means* berbasis *Python* pada data pasien Tuberkulosis di Puskesmas Simpang Kawat guna menghasilkan kluster risiko pasien (rendah, sedang, dan tinggi) sebagai dasar penyusunan strategi intervensi dan prioritas pelayanan.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis *data mining* dengan metode *Fuzzy C-Means* sebagai teknik klusterisasi. Model konseptual penelitian dirancang untuk menggambarkan alur proses penelitian secara menyeluruh mulai dari *input data* hingga menghasilkan *output* analisis klusterisasi. Model konseptual penelitian terdiri dari empat tahapan utama, yaitu: *Input Data*, *Preprocessing Data*, Proses Klusterisasi *Fuzzy C-Means*, dan *Output & Interpretasi Hasil*.

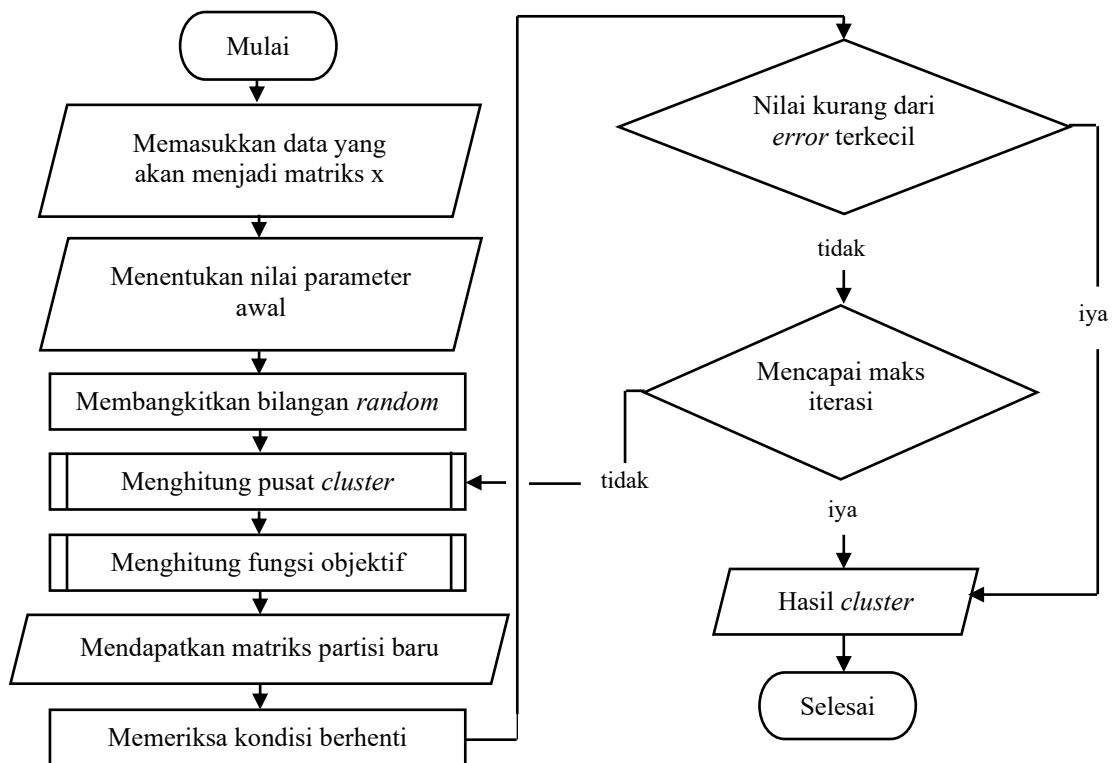
Tahap pertama adalah pengumpulan data pasien Tuberkulosis dari Puskesmas Simpang Kawat periode 2023–2024. Variabel yang digunakan terdiri dari delapan indikator klinis dan administratif, yaitu umur, jenis kelamin, klasifikasi pengobatan sebelumnya, terduga TBC RO, hasil foto toraks, hasil pemeriksaan tuberkulin/IGRA, status HIV, dan riwayat diabetes.

Tahap kedua adalah preprocessing data yang meliputi pembersihan data, transformasi ke bentuk numerik, dan normalisasi menggunakan metode *Min-Max* agar seluruh variabel berada pada skala yang seragam.

Tahap ketiga adalah penerapan metode *Fuzzy C-Means* menggunakan perhitungan manual pada *Microsoft Excel* dan implementasi kode pada *Python* sebagai pembanding. Proses klusterisasi termasuk inisialisasi parameter awal, pembentukan matriks partisi, perhitungan pusat cluster, fungsi objektif, dan iterasi hingga mencapai konvergensi.

Tahap keempat adalah interpretasi hasil yaitu pengelompokan pasien Tuberkulosis ke dalam tiga cluster risiko: rendah, sedang, dan tinggi. Hasil akhir dari metode ini digunakan untuk tujuan pendukung keputusan bagi pihak puskesmas dalam penentuan prioritas pelayanan dan intervensi.

Model konseptual penelitian menunjukkan bahwa metode *Fuzzy C-Means* bukan hanya berfungsi sebagai algoritma pengelompokan, tetapi sebagai proses analitik yang terstruktur dari *input data* hingga dihasilkan *output* berupa rekomendasi *cluster* risiko. Tahapan metode *Fuzzy C-Means* dapat dilihat pada *flowchart* berikut.



Gambar 1. Flowchart Metode Fuzzy C-Means

Alur algoritma *Fuzzy C-Means* pada *flowhart* di atas juga dapat dijelaskan dengan cara sebagai berikut:

1. Memasukan data yang akan dikelompokkan (x), dalam bentuk matriks dengan ukuran $n \times m$ yang mana n ialah jumlah sampel data, dan m sebagai variabel untuk setiap data, x_{ij} merupakan sampel data ke- i ($i = 1, 2, 3, \dots$) dan variabel ke- j ($j = 1, 2, 3, \dots$).
2. Menentukan nilai parameter awal, yaitu menginisiasi jumlah *cluster*, nilai w , maks iterasi, *error* terkecil, fungsi objektif awal dan juga iterasi awal.
3. Membangkitkan bilangan *random* (μ_{ik}) dengan $i = 1, 2, 3, \dots$; $k = 1, 2, 3, \dots$; menjadi U .
4. Menghitung pusat *cluster* ke- k (v_{kj}) dimana $k = 1, 2, 3, \dots$, dan $j = 1, 2, 3, \dots$, dengan rumus sebagai berikut:

$$v_{kj} = \frac{\sum_{i=1}^n u_{ij}^m \cdot x_i}{\sum_{i=1}^n u_{ij}^m} \quad (1)$$

5. Menghitung fungsi objektif dari iterasi ke- t (P_t), dengan rumus sebagai berikut:

$$J_m = \sum_{i=0}^n \sum_{j=0}^c u_{ij}^m \cdot \|x_i - v_j\|^2 \quad (2)$$

6. Mendapatkan matriks partisi baru/perubahan matriks partisi dengan rumus sebagai berikut:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}}} \quad (3)$$

7. Menjalankan pemeriksaan kondisi berhenti dengan ketentuan kondisi akan berhenti jika $(|P_t - P_{t-1}| < \epsilon)$ atau $(t > \text{MaxIter})$, jika tidak berhenti maka akan mengulangi Langkah 4.

Tujuan utama penelitian ini adalah mengelompokkan pasien Tuberkulosis berdasarkan karakteristik tertentu sehingga hasilnya dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan di Puskesmas Simpang Kawat. Data yang digunakan dalam penelitian ini merupakan data pasien Tuberkulosis yang didapat dari Puskesmas Simpang Kawat Kota Jambi periode 2023-2024. Data pasien Tuberkulosis yang digunakan untuk penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Data Pasien Tuberkulosis

No	Nama Lengkap Terduga TBC	Umur (Tahun)	Jenis Kelamin	Klasifikasi Berdasarkan Riwayat Pengobatan Sebelumnya	Kriteria Terduga TBC RO	Hasil Pemeriksaan Foto Toraks	Hasil Pemeriksaan Tuberkulin/IGRA	Status HIV	Riwayat Diabetes Melitus
1.	AFIQA FATYYA	8	P	Baru	Tidak	TDL	Negatif	Tidak	Tidak
2.	AMAN.MA	61	L	Baru	Tidak	TDL	Negatif	Tidak	Ya
3.	AMRIZAL	53	L	Baru	Tidak	TDL	Negatif	Tidak	Tidak
4.	ANDRI TIO PUTRA	24	L	Baru	Tidak	TDL	Negatif	ODHIV	Tidak
5.	ANGELY CLARISSA THE	24	P	Baru	Tidak	TDL	Negatif	Tidak	Tidak
....									
164.	YUSNIARTI	54	P	Baru	Tidak	TDL	Negatif	Tidak	Ya
165.	ZUHERDI	60	L	Kambuh	Pasien TB kasus kambuh (relaps), kategori 1 dan kategori 2	Negatif	Negatif	Tidak	Tidak

Sumber: Puskesmas Simpang Kawat Kota Jambi (2023-2024)

Tahapan penelitian meliputi beberapa langkah. Pertama, pengumpulan data dari puskesmas. Kedua, melakukan pra-pemrosesan data berupa pembersihan data, normalisasi agar berada pada rentang tertentu, serta penanganan data hilang. Ketiga, menentukan parameter awal algoritma, seperti jumlah *cluster*, parameter pembobot $m=2$, nilai *error threshold*, dan iterasi maksimum. Keempat, menerapkan algoritma *Fuzzy C-Means* hingga tercapai konvergensi. Hasil klasterisasi diinterpretasikan ke dalam tiga kelompok risiko, yaitu rendah, sedang, dan tinggi. Seluruh proses penelitian ini dilakukan dengan perhitungan manual pada Microsoft Excel serta bantuan *tools* Jupyter Notebook dan bahasa pemrograman *Python*.

3. Hasil dan Pembahasan

3.1. Hasil Klasterisasi *Fuzzy C-Means*

Metode *Fuzzy C-Means* diterapkan untuk mengelompokkan 165 data pasien Tuberkulosis di Puskesmas Simpang Kawat menjadi tiga *cluster*, yaitu *Cluster* Rendah, *Cluster* Sedang, dan *Cluster* Tinggi. Proses klasterisasi dilakukan melalui dua pendekatan, yaitu perhitungan manual menggunakan *Microsoft Excel* dan implementasi program *Python*. Kedua metode menunjukkan hasil yang konsisten, dengan proses konvergensi dicapai pada iterasi ke-39.

Hasil klasterisasi menunjukkan distribusi pasien sebagai berikut: *Cluster* Rendah sebanyak 63 pasien, *Cluster* Sedang sebanyak 67 pasien, dan *Cluster* Tinggi sebanyak 35 pasien. Konsistensi hasil antara perhitungan manual dan *Python* mengindikasikan bahwa parameter yang digunakan telah tepat dan stabil dalam membentuk *cluster*.

3.2. Hasil Program *Python*

Program akan menampilkan banyaknya iterasi yang telah didapatkan dari perhitungan *Fuzzy C-Means*. Banyaknya iterasi yang didapatkan adalah sama dengan perhitungan manual yaitu 39 iterasi. *Output*-nya juga menampilkan *centroid* dan menampilkan nilai fungsi objektif.

```

Iterasi 39:
Langkah 1: Menghitung centroid tiap cluster
  Fitur1  Fitur2  Fitur3  Fitur4  Fitur5  Fitur6  Fitur7  \
0  0.680029  0.874364  0.031277  0.044543  0.896152  0.088068  0.233144
1  0.699415  0.768157  0.476423  0.766909  0.897806  0.056098  0.067277
2  0.679825  0.108719  0.026056  0.041460  0.934906  0.051798  0.050338

  Fitur8
0  0.162412
1  0.098673
2  0.149838
Langkah 2: Menghitung jarak data ke tiap centroid
Langkah 3: Fungsi objektif = 56.572670
Langkah 4: Selisih fungsi objektif = 0.000009
Konvergensi tercapai pada iterasi 39
  
```

Gambar 2. Iterasi Program *Python*

Pada iterasi terakhir, akan menampilkan pula nilai selisih fungsi objektif yang didapatkan dan pada iterasi ke berapa nilai konvergensi yang tercapai, dalam hal ini adalah 39.

Centroid akhir yang telah didapat sebelumnya pada iterasi terakhir yaitu iterasi 39, ditampilkan lagi secara terpisah. Dapat dilihat pada Gambar 3.

Centroid Akhir:			
Umur	Jenis_Kelamin	\	
0	0.680029	0.874364	
1	0.699415	0.768157	
2	0.679825	0.108719	
Klasifikasi_Berdasarkan_Riwayat_Pengobatan_Sebelumnya			
Terduga_TBC_RO	\		
0	0.031277	0.044543	
1	0.476423	0.766909	
2	0.026056	0.041460	
Hasil_Pemeriksaan_Foto_Toraks	Hasil_Pemeriksaan_Tuberkulin	Riwayat_HIV	
0	0.896152	0.088068	0.233144
1	0.897806	0.056098	0.067277
2	0.934906	0.051798	0.050338
Riwayat_Diabetes			
0	0.162412		
1	0.008673		

Gambar 3. Centroid Akhir Program Python

Untuk mengetahui data ke berapa saja yang masuk ke setiap *cluster*-nya, maka program akan menampilkan data ke berapa saja yang masuk ke C1 (Sedang), C2 (Tinggi) dan C3 (Rendah). Serta akan menampilkan jumlah data di masing-masing *cluster*-nya. Dapat dilihat pada Gambar 4.

```

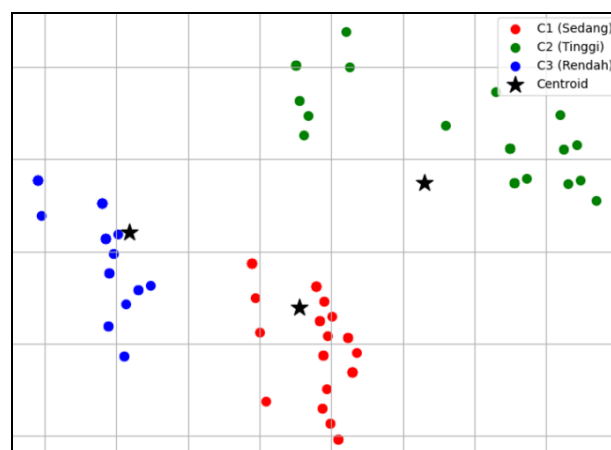
Data ke-berapa saja yang termasuk tiap cluster:
► C1 (Sedang): data ke-2, 3, 4, 6, 7, 8, 10, 11, 12, 17, 18, 20, 23, 25, 26, 30, 31, 33, 40, 41, 42, 47, 49, 51, 52, 53, 56,
59, 64, 65, 68, 69, 78, 79, 83, 86, 89, 90, 91, 93, 95, 96, 100, 101, 102, 103, 104, 105, 106, 112, 113, 114, 115, 117, 118,
122, 127, 128, 129, 133, 141, 142, 150, 156, 157, 158, 159
► C2 (Tinggi): data ke-13, 14, 15, 21, 24, 29, 32, 35, 36, 37, 44, 50, 54, 55, 57, 60, 61, 62, 63, 67, 71, 81, 92, 94, 99, 1
10, 111, 116, 131, 140, 143, 149, 153, 164, 165
► C3 (Rendah): data ke-1, 5, 9, 16, 19, 22, 27, 28, 34, 38, 39, 43, 45, 46, 48, 58, 66, 70, 72, 73, 74, 75, 76, 77, 80, 82,
84, 85, 87, 88, 97, 98, 107, 108, 109, 119, 120, 121, 123, 124, 125, 126, 130, 132, 134, 135, 136, 137, 138, 139, 144, 145, 1
46, 147, 148, 151, 152, 154, 155, 160, 161, 162, 163

Jumlah data per cluster:
C1 (Sedang): 67 data
C2 (Tinggi): 35 data
C3 (Rendah): 63 data

```

Gambar 4. Pembagian Data Setiap Cluster

Untuk dapat melihat penyebaran data, maka dibutuhkan visualisasi grafik berupa *scatter plot*. Visualisasi tersebut dapat dilihat pada Gambar 5.



Gambar 5. Visualisasi

Pada grafik visualisasi di atas, dapat dilihat bahwa penyebaran datanya cukup baik dengan bintang sebagai penanda *Centroid* atau Pusat *Cluster*-nya. Dengan visualisasi ini juga dapat dikatakan bahwa metode *Fuzzy C-Means* dapat bekerja dengan baik dalam pengelompokan data.

3.3. Hasil Akhir

Hasil akhir berupa klasterisasi data pasien Tuberkulosis yang berjumlah 165 data dan dibagi menjadi 3 *Cluster*, yaitu *Cluster* Rendah, *Cluster* Sedang dan *Cluster* Tinggi. Hasil klasterisasi dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Cluster*

<i>Cluster</i>					Keterangan
	Manual	Jumlah Data	<i>Python</i>	Jumlah Data	
Sedang	C1	67	C1	67	Sesuai
Tinggi	C2	35	C2	35	Sesuai
Rendah	C3	63	C3	63	Sesuai

Pada perhitungan manual dan program *Python*, perhitungan berhenti di iterasi yang sama, yaitu iterasi ke 39. Data-data digolongkan ke dalam masing-masing *cluster*, baik itu Rendah, Sedang dan Tinggi, memiliki beberapa alasan seperti pada Tabel 3.

Tabel 3. Pengelompokkan Data

<i>Cluster</i>	Contoh Data yg Termasuk	Alasan
Sedang	Amrizal	- Rata-rata memiliki usia yang cukup tua - Pemeriksaan belum terlalu berat/negatif - Klasifikasi baru - Tidak ada HIV/DM - Tapi memerlukan perhatian lebih lanjut.
Tinggi	Erni	- Rata-rata terdapat penyakit penyerta seperti HIV positif atau Diabetes Melitus - Pasien dengan riwayat kambuh atau relaps setelah pengobatan - Hasil pemeriksaan TBC positif dan/atau ODHIV - Memiliki usia yang relatif tua, tetapi tidak menutupi dapat terjadi pada berbagai kelompok umur
Rendah	Afiqa Fatyya	- Rata-rata usia anak-anak atau dewasa muda yang cukup sehat - hasil pemeriksaan lebih menunjukkan TBC aktif dengan hasil yang negatif - Klasifikasi sebagai kasus baru dan hasil pemeriksaan awal menunjukkan status baik

Dari 165 data, data yang terbagi di masing-masing *cluster*-nya juga sama, baik itu pada perhitungan manual ataupun program *Python*. Dapat dilihat pada Tabel 4.

Tabel 4. Data pada Setiap *Cluster*

<i>Cluster</i>		Data Ke-	Jumlah Total Data
Sedang	C1	2, 3, 4, 6, 7, 8, 10, 11, 12, 17, 18, 20, 23, 25, 26, 30, 31, 33, 40, 41, 42, 47, 49, 51, 52, 53, 56, 59, 64, 65, 68, 69, 78, 79, 83, 86, 89, 90, 91, 93, 95, 96, 100, 101, 102, 103, 104, 105, 106, 112, 113, 114, 115, 117, 118, 122, 127, 128, 129, 133, 141, 142, 150, 156, 157, 158, 159.	67
Tinggi	C2	13, 14, 15, 21, 24, 29, 32, 35, 36, 37, 44, 50, 54, 55, 57, 60, 61, 62, 63, 67, 71, 81, 92, 94, 99, 110, 111, 116, 131, 140, 143, 149, 153, 164, 165.	35
Rendah	C3	1, 5, 9, 16, 19, 22, 27, 28, 34, 38, 39, 43, 45, 46, 48, 58, 66, 70, 72, 73, 74, 75, 76, 77, 80, 82, 84, 85, 87, 88, 97, 98, 107, 108, 109, 119, 120,	63

	121, 123, 124, 125, 126, 130, 132, 134, 135, 136, 137, 138, 139, 144, 145, 146, 147, 148, 151, 152, 154, 155, 160, 161, 162, 163.	
--	---	--

3.4. Interpretasi dan Analisis Hasil Klasterisasi

Berdasarkan nilai *centroid* akhir dan karakteristik pada masing-masing *cluster*, dapat diinterpretasikan bahwa:

1. *Cluster* Rendah umumnya terdiri dari pasien dengan usia anak-anak dan dewasa muda, memiliki status pemeriksaan awal yang baik, hasil foto toraks dan tuberkulin sebagian besar negatif, tanpa komorbid penyerta seperti HIV maupun diabetes. Kelompok ini menunjukkan risiko rendah terhadap progresivitas Tuberkulosis dan membutuhkan pemantauan standar.
2. *Cluster* Sedang didominasi oleh pasien dengan usia dewasa hingga lansia awal, memiliki klasifikasi kasus baru, dan tanpa komorbid signifikan. Hasil pemeriksaan masih perlu diobservasi lebih lanjut. Kelompok ini memerlukan pemantauan lanjutan, karena memiliki peluang progresivitas Tuberkulosis jika tidak ditangani konsisten.
3. *Cluster* Tinggi terdiri dari pasien dengan risiko lebih kompleks, ditandai oleh adanya komorbiditas seperti HIV atau Diabetes Mellitus, riwayat kambuh atau pernah gagal pengobatan, dan adanya indikasi terduga TB Resistensi Obat (TBC RO). Hasil pemeriksaan penunjang umumnya menunjukkan kondisi yang mengarah pada TB aktif. Kelompok ini merupakan prioritas intervensi karena berpotensi memberikan dampak penularan dan gagal terapi lebih tinggi.

Dari pola ini terlihat bahwa faktor usia, komorbiditas, dan riwayat pengobatan merupakan variabel yang paling berpengaruh terhadap pembentukan *cluster*. Hal ini sejalan dengan literatur Tuberkulosis yang menyebutkan bahwa usia produktif dan komorbiditas berperan penting dalam menentukan tingkat keparahan Tuberkulosis serta respon terapi.

3.5. Perbandingan dengan Penelitian Terdahulu

Hasil penelitian ini memiliki kesesuaian dengan beberapa penelitian lokal yang menerapkan *Fuzzy C-Means* pada kasus Tuberkulosis. Misalnya, penelitian di salah satu Puskesmas di Kota Padang juga menghasilkan tiga *cluster* risiko TB, di mana cluster tertinggi didominasi oleh pasien dengan komorbid dan riwayat pengobatan yang tidak tuntas. Kesamaan ini menunjukkan bahwa penerapan *Fuzzy C-Means* dalam konteks pelayanan kesehatan primer di Indonesia memiliki pola segmentasi yang relatif serupa, khususnya dalam mengidentifikasi kelompok pasien berisiko tinggi.

Namun, terdapat sedikit perbedaan pola jika dibandingkan dengan penelitian di salah satu rumah sakit rujukan provinsi di Indonesia, yang menemukan bahwa faktor lingkungan seperti kepadatan hunian dan status gizi memiliki pengaruh signifikan dalam pembentukan *cluster* Tuberkulosis. Berbeda dengan penelitian tersebut, penelitian ini belum memasukkan variabel gaya hidup dan sosial ekonomi sehingga cluster terbentuk terutama oleh faktor medis dan riwayat terapi.

Perbandingan tersebut menunjukkan bahwa meskipun *Fuzzy C-Means* konsisten dalam memisahkan pasien menjadi tiga tingkat risiko, komposisi faktor pembentuk *cluster* dapat bervariasi tergantung pada karakteristik fasilitas layanan kesehatan, populasi pasien, dan jenis variabel yang digunakan. Dengan demikian, penelitian ini memberikan kontribusi pada gambaran segmentasi pasien Tuberkulosis di tingkat puskesmas di Kota Jambi.

3.6. Validasi, Keterbatasan, dan Implikasi Penelitian

Validasi dilakukan dengan membandingkan hasil perhitungan manual dan program *Python*, yang menunjukkan hasil identik baik dalam jumlah iterasi maupun komposisi *cluster*. Hal ini menunjukkan bahwa parameter *Fuzzy C-Means* yang digunakan pada jumlah *cluster*, nilai pembobot ($m=2$), *threshold error*, dan maksimum iterasi telah memberikan performa pengelompokan yang stabil.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, variabel yang digunakan masih terbatas pada data medis klinis tanpa mempertimbangkan faktor lingkungan, status gizi, sosial ekonomi, serta kepatuhan terapi yang dapat memengaruhi

perkembangan Tuberkulosis. Kedua, penelitian belum menggunakan indeks validasi *cluster* seperti Xie-Beni Index atau Dunn Index yang dapat menilai kualitas pemisahan antar *cluster* secara kuantitatif. Ketiga, ukuran sampel masih pada lingkup satu puskesmas sehingga generalisasi terhadap populasi yang lebih luas masih terbatas.

Hasil klasterisasi ini dapat digunakan oleh Puskesmas Simpang Kawat untuk memetakan prioritas penanganan pasien Tuberkulosis. *Cluster* Tinggi dapat menjadi sasaran intervensi lebih intensif seperti pemantauan ketat, intensifikasi pengobatan, dan konseling untuk mencegah resistensi obat.

4. Kesimpulan

Berdasarkan hasil penelitian mengenai klasterisasi pasien Tuberkulosis menggunakan metode *Fuzzy C-Means* (FCM), dapat disimpulkan bahwa metode FCM mampu mengelompokkan pasien ke dalam tiga kategori risiko yaitu Rendah, Sedang, dan Tinggi dengan menggunakan delapan variabel (Umur, Jenis Kelamin, Klasifikasi Berdasarkan Riwayat Pengobatan Sebelumnya, Terduga TBC RO, Hasil Pemeriksaan Foto Toraks, Hasil Pemeriksaan Tuberkulin, Riwayat HIV, dan Riwayat Diabetes Melitus). Proses perhitungan baik secara manual maupun menggunakan *Python* menunjukkan konsistensi hasil dengan konvergensi pada iterasi ke-39 dan menghasilkan pembagian *cluster* yang sama, yaitu *Cluster* Rendah (63 pasien), *Cluster* Sedang (67 pasien), dan *Cluster* Tinggi (35 pasien). Hasil ini dapat membantu Puskesmas Simpang Kawat dalam penentuan prioritas penanganan, perencanaan intervensi medis, dan alokasi sumber daya kesehatan secara lebih efektif.

Namun, penelitian ini memiliki beberapa keterbatasan, antara lain:

1. Data yang digunakan masih terbatas pada 165 pasien dari satu puskesmas, sehingga hasil klasterisasi belum mewakili kondisi populasi yang lebih luas.
2. Parameter FCM seperti jumlah *cluster*, nilai pembobot (m), dan *error threshold* masih ditentukan secara statis tanpa dilakukan optimasi.
3. Penelitian belum melakukan validasi eksternal menggunakan metode evaluasi *cluster* seperti Dunn Index, Silhouette Coefficient, atau perbandingan dengan algoritma lain.

Berdasarkan keterbatasan tersebut, disarankan pada penelitian selanjutnya untuk:

1. Memperluas jumlah dan variasi dataset, misalnya melibatkan data dari beberapa puskesmas atau rumah sakit agar hasil lebih general dan akurat.
2. Melakukan optimasi parameter FCM menggunakan metode tertentu seperti Grid Search, Particle Swarm Optimization (PSO), atau Genetic Algorithm (GA) untuk memperoleh hasil klasterisasi yang lebih optimal.
3. Menambahkan validasi dan perbandingan dengan metode lain seperti K-Means, Hierarchical Clustering, atau FCM yang terintegrasi dengan algoritma hibrida untuk meningkatkan performa klasterisasi.
4. Mengembangkan sistem analisis berbasis aplikasi atau dashboard agar hasil klasterisasi dapat digunakan secara real-time oleh tenaga medis dalam pengambilan keputusan.

Daftar Pustaka

- [1] C. Khairunnisa, C. S. Nadira and Yuziani, "TINGKAT PENGETAHUAN TATALAKSANA TUBERKULOSIS PADA MAHASISWA KEDOKTERAN UNIVERSITAS MALIKUSSALEH," *Jurnal Kedokteran dan Kesehatan Malikussaleh*, vol. 7, no. 1, pp. 27-36, 2021.
- [2] B. Yanti, "PENYULUHAN PENCEGAHAN PENYAKIT TUBERKULOSIS (TBC) ERA NEW NORMAL," *Jurnal Pengabdian Masyarakat*, vol. 4, no. 1, pp. 325-332, 2021.
- [3] D. Andriani, Sukardin, R. Ramli and N. Ilmi, "Pengetahuan dan Sikap Keluarga Dengan Pencegahan Penularan Penyakit Tuberculosis (TBC) Di Wilayah Kerja Puskesmas Penana'e Kota Bima," *Jurnal Ilmiah Ilmu Keperawatan Indonesia*, vol. 10, no. 03, pp. 109-117, 2020.
- [4] Sunarmi and Kurniawaty, "HUBUNGAN KARAKTERISTIK PASIEN TB PARU

- DENGAN KEJADIAN TUBERKULOSIS," *Jurnal 'Aisyiyah Medika*, vol. 7, no. 2, pp. 182-187, 2022.
- [5] S. Setyaningtyas, B. I. Nugroho and Z. Arif, "TINJAUAN PUSTAKA SISTEMATIS PADA DATAMINING: STUDI KASUS ALGORITMA K-MEANS CLUSTERING," *Jurnal TEKNOIF*, vol. 10, no. 2, pp. 52-61, 2022.
- [6] S. S. Prasetyo, Mustafid and A. R. Hakim, "PENERAPAN FUZZY C-MEANS KLUSTER UNTUK SEGMENTASI PELANGGAN E-COMMERCE DENGAN METODE RECENCY FREQUENCY MONETARY (RFM)," *JURNAL GAUSSIAN*, vol. 9, no. 4, pp. 421-433, 2020.
- [7] Y. D. Pradana, D. Erwanto and T. Handayani, "Implementasi Histogram Warna RGB dan Fuzzy C-Means untuk Prediksi Kebutuhan Pupuk Nitrogen Tanaman Padi," *Jurnal Ilmiah Setrum*, vol. 10, no. 2, pp. 1-12, 2021.
- [8] Aqsa and A. Supriyanto, "Fuzzy C-Means Method for Lung Image Segmentation," *Journal of Physics: Conference Series*, vol. 1899, no. 1, pp. 1–7, 2021.
- [9] Sari, D. Darwis, and M. R. Lubis, "Clustering of Hypertension Patient Data Using Fuzzy C-Means," *International Journal of Health Science and Technology*, vol. 4, no. 3, pp. 215–224, 2022.
- [10] Rahmawati, R. Meiliana, and D. Pratama, "Analisis Faktor Risiko Kejadian Tuberkulosis Paru di Wilayah Kerja Puskesmas Menggunakan Metode K-Means," *Jurnal Kesehatan Masyarakat (e-Journal)*, vol. 11, no. 4, pp. 345–354, 2023.