

Pengelompokan Penyakit Tuberkulosis Paru Berdasarkan Penyebabnya Menggunakan Metode *Clustering* (Studi Kasus : UPT Puskesmas Selesai)

Cinta Apriliza^{1*}, Relita Buaton², Hermansyah Sembiring³

¹⁻³ Sekolah Tinggi Manajemen Informatika dan Komputer Kaputama, Indonesia

Email: cintaapriliza@gmail.com¹, bbcbuaton@gmail.com², hermansyahsembiring240165@gmail.com³

Alamat: Jl. Veteran No.4A, Tangsi, Kec. Binjai Kota, Kota Binjai, Sumatera Utara 20714

Korespondensi penulis: cintaapriliza@gmail.com*

Abstract. *Pulmonary tuberculosis remains a pressing public health problem, particularly in the work area of the Duduk Health Center (UPT Puskesmas). Effective management of this disease requires a thorough understanding of the characteristics of the causes of pulmonary TB in patients. This study aims to classify pulmonary TB cases based on the main causes such as diabetes mellitus, irritant factors, pleural effusion, and family environmental conditions. The research method used is a clustering technique with the K-Means algorithm. The data used are data on pulmonary TB patients in 2020–2025 with variables of age, gender, and causative factors collected from medical records. The analysis process was carried out using MATLAB R2014b software. The clustering model was carried out in 3, 4, and 5 clusters to compare the level of segmentation efficiency. Based on the calculation results, the model with 5 clusters showed the lowest cluster variance value of 0.4889 compared to the 3-cluster model (0.7333) and 4-cluster models (0.6151), which indicates that the division into 5 clusters produces the most compact and representative data group. Each cluster shows a different combination of characteristics of pulmonary TB patients, for example: (1) elderly male patients with comorbid diabetes; (2) adolescent females with the negative influence of environmental factors; (3) adult males exposed to irritants; (4) patients with pleural effusion; and (5) groups with multiple factors. The results of this study can provide strategic input for the Finished Community Health Center UPT in formulating more targeted and targeted intervention policies in order to prevent, control, and handle pulmonary tuberculosis cases in a sustainable and effective manner.*

Keywords: *Clustering, Disease, Grouping, K-Means, Pulmonary Tuberculosis.*

Abstrak. Tuberkulosis paru masih menjadi salah satu permasalahan kesehatan masyarakat yang mendesak, khususnya di wilayah kerja UPT Puskesmas Selesai. Penanganan yang efektif terhadap penyakit ini menuntut adanya pemahaman yang mendalam mengenai karakteristik penyebab TB paru pada pasien. Penelitian ini bertujuan untuk mengelompokkan kasus TB paru berdasarkan penyebab utama seperti penyakit gula (diabetes mellitus), faktor iritan, efusi pleura, serta kondisi lingkungan keluarga. Metode penelitian yang digunakan adalah teknik clustering dengan algoritma K-Means. Data yang digunakan merupakan data pasien TB paru pada tahun 2020–2025 dengan variabel usia, jenis kelamin, dan faktor penyebab yang telah dihimpun dari rekam medis. Proses analisis dilakukan menggunakan perangkat lunak MATLAB R2014b. Model pengelompokan dilakukan dalam 3, 4, dan 5 cluster untuk membandingkan tingkat efisiensi segmentasi. Berdasarkan hasil perhitungan, model dengan 5 cluster menunjukkan nilai cluster variance terendah yaitu 0,4889 dibandingkan dengan model 3 cluster (0,7333) dan 4 cluster (0,6151), yang mengindikasikan bahwa pembagian dalam 5 cluster menghasilkan kelompok data yang paling kompak dan representatif. Setiap cluster menunjukkan kombinasi karakteristik pasien TB paru yang berbeda, misalnya: (1) pasien lansia laki-laki dengan komorbiditas penyakit gula; (2) remaja perempuan dengan pengaruh buruk faktor lingkungan; (3) dewasa laki-laki yang terpapar iritan; (4) pasien dengan efusi pleura; dan (5) kelompok dengan faktor ganda. Hasil penelitian ini dapat memberikan masukan strategis bagi UPT Puskesmas Selesai dalam menyusun kebijakan intervensi yang lebih terarah dan tepat sasaran dalam rangka pencegahan, pengendalian, dan penanganan kasus tuberkulosis paru secara berkelanjutan dan efektif.

Kata kunci: *Clustering, Penyakit, K-Means, Pengelompokan, Tuberkulosis paru,*

1. LATAR BELAKANG

Tuberkulosis paru (TB Paru) merupakan penyakit infeksi menular yang masih menjadi tantangan kesehatan masyarakat di Indonesia. Penyakit ini disebabkan oleh *Mycobacterium tuberculosis* dan lebih rentan menyerang individu dengan daya tahan tubuh lemah, seperti penderita HIV, diabetes, atau mereka yang hidup dalam kondisi lingkungan tidak sehat. Di wilayah kerja UPT Puskesmas Selesai, kasus TB Paru terus menunjukkan peningkatan setiap tahunnya, yang dipengaruhi oleh faktor lingkungan keluarga, perilaku hidup, serta rendahnya pemahaman masyarakat tentang pentingnya pengobatan dan pencegahan.

Berbagai studi sebelumnya telah menggunakan metode *K-Means Clustering* untuk menganalisis persebaran TB, seperti pada wilayah Provinsi Jawa Barat (Sulistiyono Hidayat et al., 2022) dan Kabupaten Jember (Purwanto et al., 2022), yang mengelompokkan daerah berdasarkan tingkat penyebaran kasus. Selain itu, (Okta Jaya Harmaja, 2023) menerapkan metode serupa untuk pengelompokan pasien berdasarkan jenis penyakit guna peningkatan pelayanan. Namun, sebagian besar penelitian tersebut masih berfokus pada aspek geografis atau jumlah kasus secara agregat.

Penelitian ini mengisi celah tersebut dengan menitikberatkan pada analisis penyebab TB Paru berdasarkan karakteristik individu pasien. Pendekatan ini memungkinkan identifikasi pola penyebab seperti penyakit penyerta (diabetes, efusipleura), faktor iritan, hingga lingkungan keluarga. Hal ini memberikan kebaruan dalam penggunaan metode *clustering*, karena hasilnya dapat dimanfaatkan untuk pengambilan keputusan yang lebih personal, tepat sasaran, dan kontekstual dalam upaya pengendalian TB.

Tujuan dari penelitian ini adalah untuk mengelompokkan kasus TB Paru di UPT Puskesmas Selesai berdasarkan faktor penyebabnya dengan menggunakan metode *K-Means Clustering*. Selain itu, penelitian ini juga bertujuan untuk menganalisis hubungan antara faktor penyebab dan karakteristik pasien, serta memberikan rekomendasi strategi pencegahan dan penanganan yang lebih efektif, baik secara medis maupun edukatif.

2. KAJIAN TEORITIS

Tuberkulosis

Tuberkulosis (TBC) merupakan penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*, yang umumnya menyerang organ paru-paru, tetapi juga dapat menginfeksi organ lain melalui saluran pernapasan, pencernaan, atau luka terbuka pada kulit. Gejala umum TBC antara lain batuk berdahak selama dua minggu atau lebih, sesak napas, penurunan berat badan, dan batuk berdarah (Avisa et al., 2025).

Jenis TBC yang paling banyak ditemukan adalah tuberkulosis paru (TB paru), namun TBC juga dapat menyerang organ di luar paru-paru (tuberkulosis ekstra paru), seperti otak (meningitis TB), kelenjar getah bening, pleura, kulit, tulang, payudara, dan saluran urogenital (Prameswaty & Swari, 2024). TB paru menjadi fokus utama dalam penelitian ini karena prevalensinya yang tinggi dan tingkat penularannya yang lebih mudah dibandingkan jenis TBC lainnya.

Tuberkulosis Paru

Tuberkulosis (TB) paru merupakan penyakit menular yang disebabkan oleh *Mycobacterium tuberculosis*, bakteri yang termasuk dalam kelompok Bakteri Tahan Asam (BTA). Penularan utamanya terjadi melalui droplet yang dikeluarkan saat penderita BTA positif batuk atau bersin. TB paru sangat mudah menyebar, terutama di lingkungan padat penduduk dengan sirkulasi udara yang buruk, dan meskipun utamanya menyerang paru-paru, infeksi juga dapat meluas ke organ lain seperti ginjal, tulang, kulit, dan otak (Aja et al., 2022). Beberapa faktor yang dapat meningkatkan risiko terjadinya TB paru antara lain:

- a. **Penyakit Gula** : Penyakit gula (Diabetes Mellitus) adalah gangguan metabolik kronis yang ditandai oleh kadar glukosa darah tinggi akibat gangguan insulin. Kondisi ini melemahkan sistem imun, sehingga penderita diabetes lebih rentan terinfeksi *Mycobacterium tuberculosis*. Risiko TB paru meningkat terutama pada pasien dengan kontrol gula darah yang buruk (Aditya Nugraha et al., 2021).
- b. **Faktor Iritan** : Iritan seperti asap rokok dan polusi udara dapat merusak saluran pernapasan dan melemahkan sistem imun. Kebiasaan merokok, terutama di dalam rumah, memperburuk kualitas udara dan meningkatkan risiko terinfeksi *Mycobacterium tuberculosis*, sehingga memperbesar kemungkinan terjadinya TB paru (Salsabila & R. Azizah, 2022).
- c. **EfusiPleura** : Efusi pleura adalah penumpukan cairan berlebihan di rongga pleura akibat peningkatan produksi atau penurunan penyerapan cairan. Kondisi ini dapat terjadi pada berbagai penyakit, mulai dari gangguan paru hingga inflamasi sistemik dan keganasan, serta menjadi salah satu penyebab utama gangguan pernapasan (Sari et al., 2022) .
- d. **Faktor Lingkungan Keluarga** : Ini sangat berperan dalam meningkatkan risiko penularan TB paru. Lingkungan rumah yang padat, kurang ventilasi, pencahayaan minim, dan kebiasaan merokok di dalam ruangan mempercepat penyebaran *Mycobacterium tuberculosis*, terutama jika ada anggota keluarga yang telah terinfeksi. Kontak erat seperti berbagi tempat tidur atau peralatan makan juga memperbesar

potensi penularan. Situasi ini melemahkan sistem imun dan menciptakan kondisi yang mendukung penyebaran TB paru (Afrina, 2023).

Pemahaman terhadap faktor-faktor tersebut penting sebagai dasar dalam upaya pencegahan, pengendalian, serta pengelompokan kasus TB paru berdasarkan penyebabnya.

Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini belum diketahui secara manual dari suatu kumpulan data. Data mining juga diartikan sebagai proses untuk mendapatkan informasi yang bermanfaat dari gudang basis data yang benar.

Menurut (Mai Sarah Tarigan et al., 2024) data mining merupakan sebuah konsep yang mulai berkembang sejak tahun 1990-an sebagai metode yang efektif dalam mengidentifikasi pola dan informasi yang berguna. Teknik ini digunakan untuk menemukan hubungan antar data serta mengelompokkan objek-objek ke dalam satu atau lebih *cluster*, di mana setiap objek dalam satu *cluster* memiliki tingkat kesamaan yang tinggi satu sama lain.

Clustering

Clustering adalah salah satu metode dalam data mining yang digunakan untuk menganalisis data dengan mengelompokkannya berdasarkan karakteristik tertentu. Proses ini dilakukan dengan mengukur tingkat kemiripan atau kedekatan antar objek menggunakan rumus jarak *Euclidean*.

Menurut (Ananda, 2024) *clustering* merupakan suatu metode yang digunakan untuk menemukan dan mengelompokkan data berdasarkan kesamaan karakteristik (*similarity*). *Clustering* termasuk dalam metode data mining tanpa arahan (*unsupervised*), yang berarti penerapannya tidak memerlukan proses pelatihan, bimbingan, atau target output tertentu.

Algoritma K-Means

Algoritma *K-Means* merupakan metode *clustering* non-hirarki yang digunakan untuk mengelompokkan data ke dalam sejumlah *cluster* sebanyak K, yang ditentukan terlebih dahulu. Tujuan dari algoritma ini adalah untuk mengelompokkan objek berdasarkan kemiripan atribut, sehingga data yang memiliki karakteristik serupa tergabung dalam satu kelompok, sedangkan data yang berbeda ditempatkan pada kelompok yang lain (Jelita et al., 2023).

Pengelompokkan *K-Means Clustering* dapat dilakukan dengan tahapan - tahapan sebagai berikut :

1. Menentukan jumlah *cluster* K.
2. Inisialisasi K centroid yang dapat dipilih secara acak.
3. Hitung jarak setiap data terhadap masing-masing centroid menggunakan rumus *Euclidean Distance* seperti berikut :

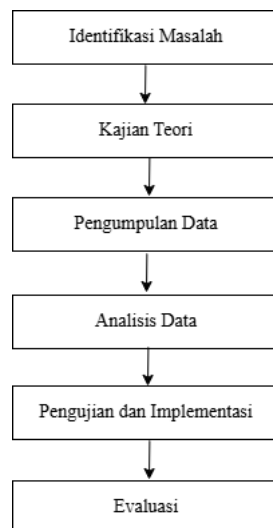
$$D_{ij} = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + (X3i - X3j)^2}$$
4. Kelompokkan data berdasarkan centroid terdekat.
5. Perbarui posisi centroid berdasarkan data dalam masing-masing *cluster*.
6. Ulangi proses mulai dari langkah 3 hingga posisi centroid tidak mengalami perubahan.

Matlab

Menurut (Fatwa et al., 2022) Matlab (*Matrix Laboratory*) adalah sebuah perangkat lunak yang menggunakan program untuk analisis dan komputasi numerik. Program ini merupakan bahasa pemrograman matematika tingkat lanjut yang dirancang dengan konsep dasar matriks. Matlab dilengkapi dengan berbagai fungsi bawaan (*built-in*) yang sangat memudahkan dalam menyelesaikan tugas-tugas kompleks dalam pengolahan angka, terutama yang berkaitan dengan perhitungan numerik berbasis matriks.

3. METODE PENELITIAN

Metodole penelitian adalah serangkaian langkah sistematis yang digunakan untuk memperoleh data dalam suatu penelitian. Dalam penelitian skripsi ini, terdapat beberapa tahapan metodologi yang diterapkan untuk mendukung proses penelitian. Tahapan tersebut ialah sebagai berikut :



Gambar 1. Struktur Metode Penelitian

Berdasarkan struktur metodologi diatas, maka berikut ini merupakan penjelasan setiap tahapannya :

1. Identifikasi Masalah

Tahapan ini merupakan tahap awal yang dilakukan untuk mengidentifikasi permasalahan yang dihadapi oleh UPT Puskesmas Selesai sebagai objek penelitian ini.

2. Kajian Teori

Tahapan ini bertujuan untuk mengumpulkan berbagai teori yang relevan dengan permasalahan yang sedang diteliti. Pada tahap ini, teori diperoleh dari berbagai sumber, seperti jurnal, artikel, dan referensi lainnya.

3. Pengumpulan Data

Tahapan ini merupakan tahap mengumpulkan data penelitian, yaitu data pasien penyakit Tuberkulosis paru berdasarkan variabel yang diperlukan dalam pembuatan skripsi ini, yang mana lokasi penelitian ini dilakukan di UPT Puskesmas Selesai.

4. Analisis Data

Pada tahap ini, data yang telah diperoleh dari proses pengumpulan sebelumnya diolah dan dianalisis, sehingga dapat diklasifikasikan sesuai dengan variabel yang telah ditetapkan dalam penelitian ini.

5. Pengujian dan Implementasi

Tahapan ini merupakan tahapan yang melakukan pengujian validasi dan implementasi data yang telah dianalisis sebelumnya.

6. Evaluasi

Tahap akhir ini bertujuan untuk merumuskan kesimpulan, yang menggambarkan hasil penelitian secara keseluruhan. Melalui saran yang diberikan, diharapkan dapat membawa perbaikan dan manfaat bagi pihak lain

Data dalam penelitian ini berasal dari UPT Puskesmas Selesai sebanyak 563 data, dengan variabel Usia, Jenis Kelamin, dan Penyebab yang berperan penting dalam mengidentifikasi pola penyebaran Tuberkulosis paru. Seluruh analisis dilakukan menggunakan MATLAB untuk pengolahan data dan penerapan algoritma *Clustering*. Karena data awal bersifat kategorikal, seperti Usia, Jenis Kelamin, dan Penyebab, maka dilakukan konversi ke bentuk numerik agar dapat diolah lebih lanjut dan menghasilkan informasi yang lebih membantu dalam pengendalian TB paru di UPT Puskesmas Selesai.

4. HASIL DAN PEMBAHASAN

Input Data

Data yang digunakan sebagai input dalam sistem berasal dari hasil penelitian yang dilakukan di UPT Puskesmas Selesai. Data tersebut telah melalui proses transformasi berdasarkan nilai-nilai yang disesuaikan untuk setiap variabel yang digunakan. Berikut adalah rincian data yang dapat dimasukkan ke dalam sistem, termasuk variabel yang digunakan beserta nilai transformasi yang diterapkan:

1. Input Data

Jumlah Data : 563 data

Variabel : X : Usia
: Y : Jenis Kelamin
: Z : Penyebab

2. Pengelompokkan *cluster* : 3 *cluster*, 4 *cluster*, 5 *cluster*

3. Nilai Transformasi

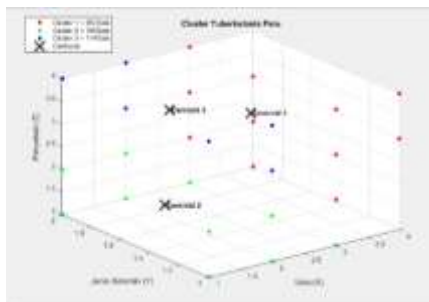
Tabel 1. Nilai Transformasi Data

Variabel	Transformasi	Nilai Transformasi	Keterangan
Usia	12 – 25 tahun	1	Remaja
	26 – 45 tahun	2	Dewasa
	46 – 65 tahun	3	Lansia
	65 - atas	4	Manula
Jenis Kelamin	P	1	Perempuan
	L	2	Laki-laki
Penyebab	Faktor Iritan	1	-
	Penyakit Efusi pluera	2	
	Penyakit Gula	3	
	Faktor Lingkungan Keluarga	4	

Hasil Clustering Menggunakan Matlab

Untuk melakukan pengelompokkan berdasarkan jarak kedekatan variabel, digunakan rumus jarak Euclidean. Pengelompokkan dilakukan dengan 3 *Cluster*, 4 *Cluster* dan 5 *Cluster*.

1. Pengelompokkan 3 *Cluster*

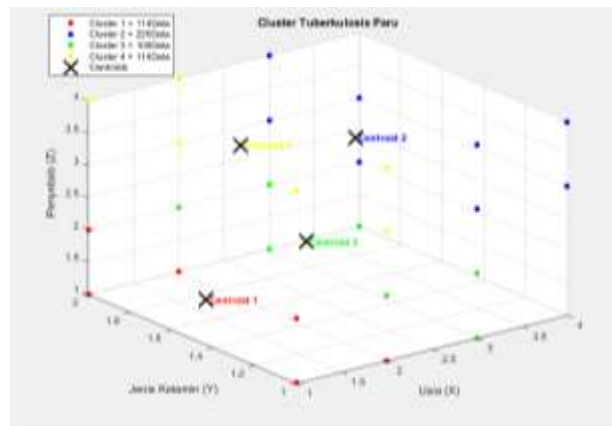


Gambar 2. Pengelompokkan 3 *Cluster*

Berdasarkan gambar diatas, dapat diketahui bahwa:

- Pada *cluster* 1 dengan jumlah data 262 data, dapat dikelompokkan berdasarkan usia “lansia” dengan jenis kelamin “laki-laki” penyebabnya adalah “penyakit gula”.
- Pada *cluster* 2 dengan jumlah data 186 data, dapat dikelompokkan berdasarkan usia “dewasa” dengan jenis kelamin “laki-laki” penyebabnya adalah “faktor iritan”.
- Pada *cluster* 3 dengan jumlah data 114 data, dapat dikelompokkan berdasarkan usia “remaja” dengan jenis kelamin “perempuan” penyebabnya adalah “faktor lingkungan keluarga”.

2. Pengelompokan 4 Cluster

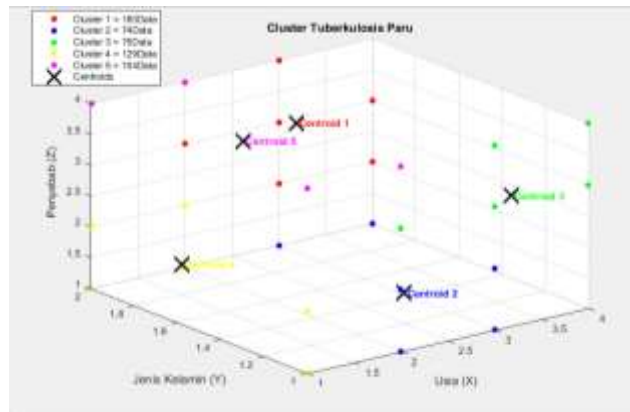


Gambar 3. Pengelompokan 4 Cluster

Berdasarkan gambar diatas, dapat diketahui bahwa:

- Pada *cluster* 1 dengan jumlah data 114 data, dapat dikelompokkan berdasarkan usia “dewasa” dengan jenis kelamin “laki-laki” penyebabnya adalah “faktor iritan”.
- Pada *cluster* 2 dengan jumlah data 225 data, dapat dikelompokkan berdasarkan usia “lansia” dengan jenis kelamin “laki-laki” penyebabnya adalah “penyakit gula”.
- Pada *cluster* 3 dengan jumlah data 109 data, dapat dikelompokkan berdasarkan usia “lansia” dengan jenis kelamin “laki-laki” penyebabnya adalah “penyakit efusipluera”.
- Pada *cluster* 4 dengan jumlah data 114 data, dapat dikelompokkan berdasarkan usia “remaja” dengan jenis kelamin “perempuan” penyebabnya adalah “faktor lingkungan keluarga”.

3. Pengelompokan 5 Cluster



Gambar 4. Pengelompokan 5 Cluster

Berdasarkan gambar diatas, dapat diketahui bahwa:

- Pada *cluster* 1 dengan jumlah data 180 data, dapat dikelompokkan berdasarkan usia “lansia” dengan jenis kelamin “laki-laki” penyebabnya adalah “penyakit gula”.
- Pada *cluster* 2 dengan jumlah data 74 data, dapat dikelompokkan berdasarkan usia “lansia” dengan jenis kelamin “perempuan” penyebabnya adalah “penyakit efusipluera”.
- Pada *cluster* 3 dengan jumlah data 75 data, dapat dikelompokkan berdasarkan usia “lansia” dengan jenis kelamin “perempuan” penyebabnya adalah “penyakit gula”.
- Pada *cluster* 4 dengan jumlah data 129 data, dapat dikelompokkan berdasarkan usia “dewasa” dengan jenis kelamin “laki-laki” penyebabnya adalah “faktor iritan”.
- Pada *cluster* 5 dengan jumlah data 104 data, dapat dikelompokkan berdasarkan usia “remaja” dengan jenis kelamin “perempuan” penyebabnya adalah “faktor lingkungan keluarga”.

Hasil Pengujian

Setelah data terbagi ke dalam beberapa *cluster*, peneliti melanjutkan dengan melakukan evaluasi terhadap hasil pengelompokan tersebut. Evaluasi ini bertujuan untuk memastikan bahwa setiap *cluster* yang terbentuk memiliki kualitas yang baik dan mampu merepresentasikan karakteristik data secara akurat. Pengujian dilakukan melalui analisis *cluster* dengan menghitung nilai *variance* atau sebaran data dalam masing-masing *cluster*. Dalam perhitungan ini digunakan tiga indikator utama, yaitu $V_{minimum}$, $V_{maksimum}$, dan *cluster variance*. *Variance* setiap *cluster* dihitung menggunakan rumus berikut:

$$Variance_k = \frac{1}{n_k} \sum_{i=0}^{n_k} (x_i - c_k)$$

Dimana:

n_k : Jumlah titik data dalam *cluster*

x_i : Titik data ke-i dalam *cluster*

c_k : Centroid dari *cluster*

1. $V_{minimum}$ (Minimum Variance): Merupakan nilai varians terkecil yang ditemukan di antara seluruh *cluster* dalam satu set data. Nilai ini menunjukkan seberapa kecil penyebaran data pada *cluster* yang paling rapat.
2. $V_{maksimum}$ (Maximum Variance): Merupakan nilai varians terbesar di antara seluruh *cluster*. Nilai ini mencerminkan seberapa besar penyebaran data dalam *cluster* yang paling tersebar.
3. *Cluster Variance*: Merupakan nilai rata-rata dari semua varians *cluster* dalam satu set data. Indikator ini digunakan untuk melihat tingkat penyebaran data secara umum di seluruh *cluster* yang terbentuk.

Rumus *cluster variance*:

$$Cluster\ Variance = \frac{1}{k} \sum_{i=1}^k variance_k$$

4. Hasil Pengujian 3 Cluster, 4 Cluster dan 5 Cluster

Tabel 2. Hasil Pengujian 3 Cluster, 4 Cluster dan 5 Cluster

Cluster	Centroid	Variance	V_{min}	V_{max}	Cluster Variance
3	3,2; 1,7; 2,9	6,2078	0,16	9,03	13,9384
	1,9; 1,7; 1,3	12,1974	0,20	12,07	
	1,4; 1,5; 3,9	23,4101	0,40	15,38	
4	1,7; 1,7; 1,1	0,4707	0,19	1,79	0,5464
	3,2; 1,7; 3,1	0,4831	0,14	1,94	
	2,5; 1,6; 1,8	0,6555	0,45	3,05	
	1,4; 1,5; 3,9	0,5761	0,42	1,42	
5	3,2; 2,0; 2,9	0,3548	0,04	1,52	0,4889
	2,5; 1,2; 1,5	0,6886	0,54	3,14	
	3,2; 1,0; 3,1	0,4089	0,04	1,46	
	1,7; 1,9; 1,3	0,5091	0,18	1,80	
	1,4; 1,5; 4,0	0,4829	0,35	0,68	

Berdasarkan tabel di atas, dapat dilihat bahwa:

1. Pengujian 3 Cluster

- a) *Cluster* pertama memiliki nilai *variance* sebesar 6,2078, yang menunjukkan tingkat penyebaran data sedang. Nilai V_{min} sebesar 0,16 dan V_{max} sebesar 9,03 menandakan adanya sebaran data yang cukup lebar.

- b) *Cluster* kedua menunjukkan *variance* sebesar 12,1974, yang berarti variasi data dalam *cluster* ini cukup tinggi. V_{min} sebesar 0,20 dan V_{max} sebesar 12,07 juga menunjukkan penyebaran data yang luas dan mengindikasikan adanya heterogenitas dalam *cluster*.
- c) *Cluster* ketiga memiliki *variance* paling tinggi yaitu 23,4101, dengan V_{min} 0,40 dan V_{max} 15,38, menunjukkan bahwa data dalam *cluster* ini sangat bervariasi dan kemungkinan terdapat data ekstrem.

Nilai *Cluster Variance* keseluruhan adalah 13,9384, yang menunjukkan tingkat penyebaran data antar *cluster* cukup tinggi dan tidak stabil.

2. Pengujian 4 Cluster

- a) *Cluster* pertama memiliki *variance* 0,4707, yang merupakan salah satu yang terendah. Dengan V_{min} 0,19 dan V_{max} 1,79, menunjukkan bahwa data cukup homogen dan tersebar secara sempit.
- b) *Cluster* kedua menunjukkan *variance* 0,4831, dengan V_{min} 0,14 dan V_{max} 1,94, yang juga mengindikasikan distribusi data yang stabil dan homogen.
- c) *Cluster* ketiga memiliki *variance* sebesar 0,6555, yang sedikit lebih tinggi dibanding dua *cluster* sebelumnya. Dengan V_{min} 0,45 dan V_{max} 3,05, ada sedikit peningkatan variasi data, namun masih dalam batas wajar.
- d) *Cluster* keempat menunjukkan *variance* 0,5761, V_{min} 0,42 dan V_{max} 1,42, juga menunjukkan sebaran data yang moderat.

Nilai *Cluster Variance* keseluruhan adalah 0,5464, yang menandakan bahwa model 4 *cluster* menghasilkan pengelompokan yang cukup kompak dan stabil.

3. Pengujian 5 Cluster

- a) *Cluster* pertama memiliki *variance* 0,3548, dengan V_{min} 0,04 dan V_{max} 1,52, menunjukkan sebaran data yang sangat rendah dan homogen.
- b) *Cluster* kedua memiliki *variance* 0,6886, V_{min} 0,54 dan V_{max} 3,14, yang menunjukkan penyebaran data lebih besar dibanding *cluster* lainnya.
- c) *Cluster* ketiga menunjukkan *variance* 0,4089, dengan V_{min} 0,04 dan V_{max} 1,46, menandakan sebaran data yang rendah dan stabil.
- d) *Cluster* keempat memiliki *variance* 0,5091, dengan V_{min} 0,18 dan V_{max} 1,80, menunjukkan sebaran data yang relatif sempit.
- e) *Cluster* kelima menunjukkan *variance* 0,4829, V_{min} 0,35 dan V_{max} 0,68, mengindikasikan *cluster* yang sangat kompak dan stabil.

Nilai *Cluster Variance* keseluruhan adalah 0,4889, yang merupakan yang paling rendah dibandingkan model 3 dan 4 *cluster*. Ini menunjukkan bahwa model dengan 5 *cluster* adalah yang paling optimal karena menghasilkan pengelompokan paling homogen dan stabil.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan terkait pengelompokan penyakit tuberkulosis paru berdasarkan penyebabnya dengan menerapkan metode *clustering* pada studi kasus di UPT Puskesmas Selesai, maka dapat disimpulkan beberapa hal sebagai berikut: Pengelompokan data berhasil dilakukan menggunakan metode *K-Means* dengan tiga variabel utama, yaitu usia, jenis kelamin, dan penyebab. Proses analisis dilakukan dengan pendekatan 3, 4, dan 5 *cluster* menggunakan perangkat lunak MATLAB R2014b. Pengelompokan dengan 3 *cluster* menghasilkan tiga kelompok utama, yakni: pasien lansia laki-laki dengan penyakit gula, pasien dewasa laki-laki dengan faktor iritan, dan pasien remaja perempuan dengan penyebab lingkungan keluarga. Model 4 *cluster* menunjukkan pola serupa dengan penambahan satu *cluster* baru, yaitu pasien lansia laki-laki yang mengalami efusi pleura, menandakan keberagaman penyebab semakin teridentifikasi seiring meningkatnya jumlah *cluster*.

Pengelompokan 5 *cluster* menghasilkan segmentasi yang lebih rinci, termasuk pemisahan pasien lansia berdasarkan jenis kelamin dengan penyebab yang sama (penyakit gula), serta penambahan *cluster* efusi pleura pada lansia perempuan. Ini menunjukkan bahwa peningkatan jumlah *cluster* mampu menangkap variasi data yang lebih halus. Secara umum, penyebab utama tuberkulosis paru dipengaruhi oleh usia dan jenis kelamin. Penyakit gula dominan pada lansia, terutama laki-laki, namun juga muncul pada perempuan. Faktor iritan paling banyak terjadi pada laki-laki usia dewasa, sementara faktor lingkungan keluarga cenderung menjadi penyebab pada remaja perempuan. Efusi pleura lebih sering ditemukan pada pasien lansia baik laki-laki maupun perempuan. Dari hasil pengujian, model 5 *cluster* terbukti paling optimal karena menghasilkan nilai *cluster variance* terendah sebesar 0,4889, yang menunjukkan bahwa data dalam masing-masing *cluster* lebih kompak, homogen, dan representatif dibandingkan model dengan 3 dan 4 *cluster*.

Berdasarkan temuan dan kesimpulan penelitian, disarankan agar penelitian selanjutnya menggunakan algoritma *clustering* selain *K-Means* dan menambahkan variabel seperti riwayat merokok atau kondisi lingkungan untuk meningkatkan akurasi pengelompokan. Selain itu, evaluasi kualitas *cluster* sebaiknya menggunakan metode alternatif seperti *Silhouette Coefficient* atau *Davies-Bouldin Index* guna memperkuat validitas hasil.

DAFTAR REFERENSI

- Aditya Nugraha, I. B., Gotera, W., & Yustin, W. E. F. (2021). Diabetes melitus sebagai faktor risiko tuberkulosis. *Jurnal Kedokteran Meditek*, 27(3), 273–281. <https://doi.org/10.36452/jkdoktmeditek.v27i3.2126>
- Afrina, Y. (2023). Faktor lingkungan dengan kejadian tuberkulosis paru. *Jurnal Riset Kesehatan Poltekkes Depkes Bandung*, 15(1), 1–21.
- Aja, N., Ramli, R., & Rahman, H. (2022). Penularan tuberkulosis paru dalam anggota keluarga di wilayah kerja Puskesmas Siko Kota Ternate. *Jurnal Kedokteran dan Kesehatan*, 18(1), 78–87. <https://doi.org/10.24853/jkk.18.1.78-87>
- Ananda, R. A. (2024). *Clustering menggunakan algoritma K-means untuk mengelompokkan data perjudian berdasarkan wilayah di Kota Binjai (Studi kasus: Pengadilan Negeri Binjai)* [Skripsi, Universitas XYZ].
- Avisa, D., Widyadhana, B., D, K. C., & Pawitra, A. S. (2025). Gambaran epidemiologi penyakit tuberkulosis wilayah kerja Dinas Kesehatan Kabupaten Kediri. *Jurnal Kesehatan Masyarakat*, 6, 3015–3022.
- Fatwa, M., Rizki, R., Sriwinarty, P., & Supriyadi, E. (2022). Pengaplikasian MATLAB pada perhitungan matriks. *Papanda Journal of Mathematics and Science Research*, 1(2), 81–93. <https://doi.org/10.56916/pjmsr.v1i2.260>
- Harmaja, O. J., Widyatama, H. H., & Suryadi, S. L. (2023). Implementasi algoritma K-means clustering untuk pengelompokan penyakit pasien pada Puskesmas Pulo Brayan. *Jurnal Sains dan Teknologi*, 5(1), 150–157.
- Hidayat, F. S., Berliana, R., Affandi, P., Zuliana, V., & Padilah, T. N. (2022). Penerapan K-means clustering dalam pengelompokan kasus tuberkulosis di Provinsi Jawa Barat. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 213–227. <https://doi.org/10.5281/zenodo.7049113>
- Jelita, T., Buaton, R., Simajuntak, M., & Kaputama, S. (2023). Pengelompokan bidang usaha terhadap bantuan produktif usaha mikro (BPUM) berdasarkan wilayah Deli Serdang menggunakan metode clustering K-means (Studi kasus: Dinas Koperasi dan UMKM Kabupaten Deli Serdang). *Journal of Computer Science and Information Technology*, 3(2), 50–60.
- Prameswaty, A. A., & Swari, M. H. P. (2024). Perancangan sistem pakar diagnosis penyakit TBC paru dengan metode Certainty Factor dan Dempster Shafer. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(5), 8658–8663.
- Purwanto, B., Nilogiri, A., & Wardoyo, A. E. (2022). Penerapan algoritma K-means clustering untuk pengelompokan penyebaran penyakit TBC (Studi kasus: Puskesmas di Kabupaten Jember). *Jurnal Smart Teknologi*, 3(3), 1–10. <http://jurnal.unmuhjember.ac.id/index.php/JST>
- Salsabila, D. S., & Azizah, R. (2022). Faktor risiko kejadian tuberkulosis paru di Indonesia. *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, 5(9), 1054–1062. <https://jurnal.unismuhpalu.ac.id/index.php/MPPKI/article/view/2622/2309>

- Sari, D. P., & Putri, A. M. (2019). Analisis spasial penderita tuberkulosis paru dengan pendekatan GIS dan K-means clustering di Kota Medan. *Jurnal Kesehatan Masyarakat*, 14(3), 243–251. <https://doi.org/10.20473/jkm.v14i3.2019.243-251>
- Sari, G. K., Sarifuddin, & Setyawati, T. (2022). Tuberkulosis paru post-Wodec pleural effusion: Laporan kasus. *Jurnal Medical Profession*, 4(2), 174–182.
- Tarigan, M. S., Hardinata, J. T., Qurniawan, H., Salfii, M., & Winanjaya, R. (2024). Implementasi data mining menggunakan algoritma Apriori dalam menentukan persediaan barang (Studi kasus: Toko Sinar Harahap). *Justify: Jurnal Sistem Informasi Ibrahimy*, 3(1), 55–65. <https://doi.org/10.35316/justify.v3i1.5335>