

DATA MINING UNTUK ANALISA TINGKAT KESEMBUHAN PASIEN TUBERKULOSIS (TB)

¹Podo Wiseso ²Muhammad Reza Redo Islami*

¹Fakultas Teknologi dan Bisnis, Universitas Dharma Wacana

²Politeknik Negeri Lampung

ABSTRAK

Penelitian ini memanfaatkan teknik data mining untuk menganalisis tingkat kesembuhan pasien tuberkulosis (TB) di Provinsi Lampung, Indonesia, dengan menggunakan data dari Sistem Informasi Tuberkulosis Komunitas (SITK) dan Row Data Individu (RDI). Tujuan utama adalah mengidentifikasi faktor risiko yang mempengaruhi penularan TB, termasuk karakteristik individu dan faktor lingkungan. Metode K-means clustering digunakan untuk mengelompokkan data berdasarkan gejala dan durasi pengobatan, menghasilkan 19 kluster dengan pola spesifik. Gejala seperti demam meriang, sesak napas, dan batuk ditemukan penting dalam menentukan tingkat kesembuhan pasien. Analisis menunjukkan variasi signifikan dalam distribusi gejala di antara kluster, dengan beberapa kluster menunjukkan respons pengobatan yang lebih baik. Penelitian ini diharapkan dapat berkontribusi dalam upaya pengendalian TB dan mencapai Tujuan Pembangunan Berkelanjutan (SDGs) terkait kesehatan.

Keyword: Analisis Data, Clustering, K-means, RDI

1. PENDAHULUAN

Kesehatan merupakan aspek fundamental dalam kehidupan individu dan masyarakat. Kesehatan yang baik tidak hanya meningkatkan kualitas hidup individu, tetapi juga berkontribusi pada perkembangan ekonomi dan sosial. Dalam konteks ini, tuberkulosis (TB) menjadi salah satu isu kesehatan global yang mendesak untuk ditangani, terutama di negara-negara berkembang, termasuk Indonesia. Dengan prevalensi yang tinggi dan munculnya resistensi obat, TB menuntut perhatian serius dari berbagai pihak, termasuk pemerintah, masyarakat, dan organisasi non-pemerintah.

Menurut laporan Global TB Report tahun 2022 (*Global Tuberculosis Report 2022*, n.d.), Indonesia menduduki peringkat kedua di dunia dengan jumlah kasus tuberkulosis terbanyak, mencapai sekitar 969 ribu kasus. Kasus TB ini terutama menyerang kelompok usia produktif, yaitu individu berusia antara 45 hingga 54 tahun. Data ini menunjukkan urgensi penanganan TB di Indonesia dan perlunya upaya yang lebih besar untuk mengendalikan penyakit ini, terutama dalam kelompok usia yang sangat berkontribusi

pada kegiatan ekonomi dan sosial. Selain itu, tantangan dalam penanganan TB juga mencakup kurangnya akses terhadap layanan kesehatan yang berkualitas, stigma sosial, dan keterbatasan sumber daya.

Penelitian ini akan menggunakan pendekatan kuantitatif dengan analisis data dari Sistem Informasi Tuberkulosis Komunitas (SITK) dan Row Data Individu (RDI). Data ini akan digunakan untuk menganalisis faktor risiko tuberkulosis, termasuk karakteristik individu dan faktor lingkungan yang berkontribusi terhadap penularan TB. Metode analisis statistik yang tepat akan diterapkan untuk mengidentifikasi hubungan antara faktor-faktor risiko dan tingkat penularan TB.

Studi-studi sebelumnya juga pernah dilakukan analisis kluster terhadap mahasiswa dari berbagai jurusan dan daerah, membaginya menjadi tiga kluster berdasarkan karakteristik mereka (Annas & Wahab, 2023) adapun beberapa penelitian yang menunjukkan bahwa pemahaman mendalam tentang faktor risiko individu dan lingkungan sangat penting dalam upaya pencegahan dan pengendalian TB.

Penelitian sebelumnya telah mengidentifikasi berbagai faktor risiko, karakteristik transaksi akun individu, yang dianggap tidak cukup untuk mengidentifikasi pelanggan pada bank secara lebih akurat (Hasheminejad & Khorrami, 2021) namun masih terdapat kekurangan dalam analisis yang lebih mendalam mengenai interaksi antara faktor-faktor tersebut (*Global Tuberculosis Report 2022*, n.d.). Dengan memanfaatkan data RDI dari SITK, penelitian ini bertujuan untuk memberikan kontribusi baru dalam literatur yang ada, serta mendukung upaya global untuk mencapai Tujuan Pembangunan Berkelanjutan (*SDGs*), khususnya *SDG* nomor 3 yang menekankan kesehatan yang baik dan kesejahteraan bagi semua.

Berdasarkan fakta-fakta yang ada, dugaan awal penelitian ini adalah bahwa terdapat hubungan signifikan antara faktor risiko individu (seperti riwayat paparan TB, status imun, dan perilaku kesehatan) dan tingkat penularan tuberkulosis di Provinsi Lampung. Selain itu, diharapkan bahwa faktor lingkungan, seperti akses terhadap layanan kesehatan dan tingkat penularan di masyarakat, juga berkontribusi terhadap peningkatan kasus TB. Dengan memahami faktor-faktor ini, penelitian ini bertujuan untuk menganalisa tingkat kesembuhan pasien TB dalam upaya mengendalikan penyebaran TB di wilayah Provinsi Lampung.

2. LITERATUR REVIEW

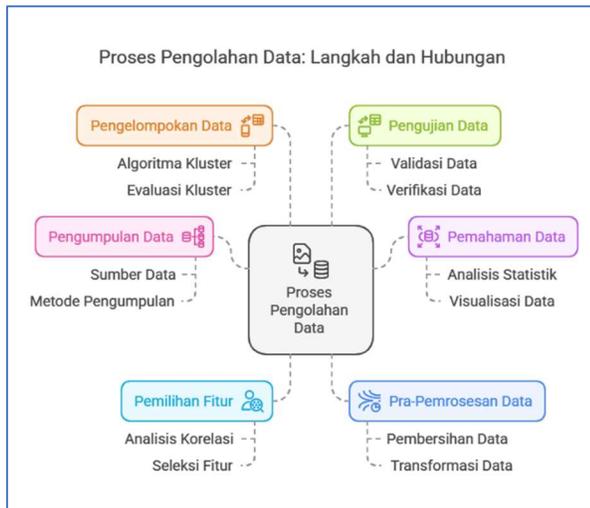
Adapun penelitian sebelumnya yang pernah dilakukan dan menjadi penguat penelitian ini diantaranya K-means (Ahmadian et al., 2024; Biswas, 2024; Harpad et al., 2024) dalam penelitian ini adalah meminimalkan jumlah semua jarak intra-kluster, yang juga disebut sebagai jumlah kesalahan kuadrat (sum of squared errors, SSE.) (Li et al., 2022) pada proses preprocessing data dilakukan transformasi data, hal ini juga pernah dilakukan pada penelitian sebelumnya dimana data preprocessing

melibatkan konversi teks menjadi huruf kecil (Abbas et al., 2024; Acheampong, 2020; Alamoodi, 2021) adapun proses datamining sendiri disini untuk mendapatkan dan mengekstrak informasi dari data yang akan di mining (Alharthi & Abdullah, 2022; Alim, 2024; Alkattan et al., 2024; Banait & SANE, 2022)

3. METODOLOGI

3.1 Pengumpulan Data

Data akan dikumpulkan dari Sistem Informasi Tuberkulosis Komunitas (*SITK*) dan *Row Data Individu (RDI)* pasien TB di Provinsi Lampung. Data ini akan mencakup informasi demografis, riwayat medis, faktor risiko, dan hasil pengobatan. Proses pengumpulan data yang dilakukan dalam penelitian ini menghasilkan *Row Data Individu (RDI)* pasien Tuberkulosis (TB) untuk periode Januari hingga Juli 2022. Tahapan pengumpulan data dimulai dengan mengajukan permohonan izin kepada **Lembaga Inisiatif Lampung Sehat (ILS)** secara langsung. Izin diperoleh secara lisan dari pihak lembaga yang bersangkutan. Setelah memperoleh izin, data diunduh melalui laman <http://sitk.online/> (*SITK Penabulu-STPI*, n.d.) yang merupakan **Sistem Informasi Tuberkulosis Komunitas (SITK)**. Data yang diperoleh berupa *RDI* pasien yang telah ternotifikasi selama periode tersebut dan disimpan dalam format *Microsoft Excel Worksheet*.



Gambar 1. Proses metode pengolahan data

3.2 Pemahaman Data

Hasil pengumpulan data berupa RDI digunakan sebagai bahan utama untuk menganalisis tingkat kesembuhan pasien TB. Pemahaman data dilakukan dengan mempelajari setiap fitur dalam dataset, termasuk tipe data, fungsi, serta korelasinya. Berikut adalah hasil pemahaman tipe data pada dataset.

Tabel 1 Data RDI

No	Nama Fitur	Tipe Data	Peran	Nilai
1	No	Numeric	Feature	1 hingga 1.726
2	Tanggal Laporan	Datetime	Feature	-
3	ID Pasien (1)	Text	Meta	-
4	NIK	Text	Meta	-
5	Nama Lengkap	Text	Meta	-
6	Tipe Pasien	Categorical	Feature	BTA+, Ekstra Paru, Rontgen+, TBC Anak+, TCM+

Dataset ini terdiri atas 1.726 instance, 44 fitur, dan 14 atribut meta. Fitur-fitur ini meliputi tipe data categorical, numeric, dan datetime, yang relevan untuk analisis tingkat kesembuhan pasien TB.

Dalam analisis awal, data menunjukkan distribusi gejala pasien TB berdasarkan hasil pengobatan (sembuh, gagal, meninggal, dsb.). Berikut adalah rekapitulasi sebaran gejala

Tabel 2. Sebaran gejala

Faktor Gejala	Jumlah Kasus	Persentase Sebaran
Batuk	1.655	96%
Sesak Nafas	1.157	67%
Berkeringat malam hari tanpa kegiatan	943	55%
Demam meriang >1 bulan	478	28%
DM	43	2%

Sebaran usia pasien TB juga dianalisis berdasarkan kelompok umur sesuai pedoman Kementerian Kesehatan RI (2009). Selain itu, durasi pengobatan pasien teridentifikasi bervariasi antara 135 hari hingga 213 hari, dengan mayoritas pasien menyelesaikan pengobatan dalam rentang waktu 160 hingga 170 hari.

Tabel 3. Sebaran Pasien

Kelompok Usia	Rentang Usia	Jumlah Kasus	Persentase
Anak-anak	1–14 tahun	153	9%
Remaja	15–25 tahun	224	13%
Dewasa	26–59 tahun	964	56%
Lansia	>60 tahun	385	22%

3.3. Pra-pemrosesan Data

Data akan diolah untuk membersihkan data yang tidak lengkap atau tidak akurat, mengatasi missing values, dan melakukan normalisasi data jika diperlukan. Tahapan pra-pemrosesan mencakup pembersihan, transformasi, dan pengorganisasian dataset untuk analisis lebih lanjut. Beberapa langkah utama yang dilakukan adalah, Transformasi tipe data fitur usia, yang semula dalam format kategorikal dengan tambahan teks "Thn", diubah menjadi format numerik untuk memudahkan analisis, Labeling lama pengobatan dan durasi pengobatan pasien yang dikategorikan menjadi tiga label berdasarkan standar pengobatan TB selama 6 bulan (182 hari) dimana kategori Cepat: 135–167 hari, Tepat Waktu: 168–183 hari dan Lama: 184–213 hari

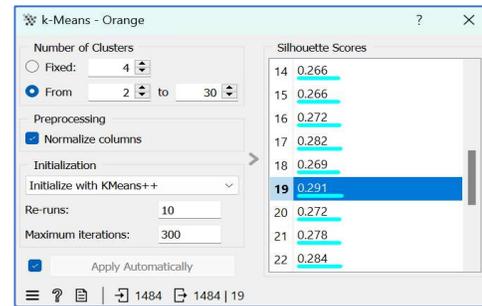
3.4 Pemilihan Fitur (Feature Selection)

Pemilihan fitur dilakukan untuk menyaring variabel yang paling relevan dengan tujuan penelitian. meliputi: Seleksi Baris (Select Rows)

dimana pada tahap ini, data disaring berdasarkan fitur Tipe Pasien dengan hasil BTA+, Ekstra Paru, Rontgen+, dan TCM Anak. Data TB Anak (usia 0–14 tahun) dikeluarkan untuk meningkatkan akurasi analisis. Setelah seleksi ini, jumlah dataset berkurang menjadi 1.574 *instance*. selanjutnya adalah melakukan Seleksi Hasil Pengobatan berdasarkan fokus penelitian, hanya hasil pengobatan dengan kategori "Sembuh" dan "Lengkap Pengobatan" yang dipertahankan. Setelah seleksi, jumlah instance berkurang menjadi 1.484. selanjutnya dilakukan seleksi Kolom (Select Columns) guna mempertahankan beberapa Fitur yang meliputi: Usia, hal ini dimasukkan Untuk menganalisis tingkat kesembuhan berdasarkan kelompok usia. selanjutnya, lama Pengobatan: Untuk mengevaluasi durasi pengobatan pasien, dan terakhir adalah Gejala TB yang berisi variabel Batuk, Sesak Nafas, Demam meriang >1 bulan, DM, dan Berkeringat malam hari tanpa kegiatan.

3.5 Pengelompokan Data (Clustering)

Tahapan selanjutnya adalah proses pengelompokan data (*clustering*) menggunakan algoritma *K-Means Clustering*, yang bertujuan untuk mengelompokkan data ke dalam klaster berdasarkan pola dan karakteristik yang ada. Kmeans sendiri dirancang untuk klasterisasi data streaming melalui proses multi-fase yang melibatkan pembuatan mikro-klaster dan pembentukan makro-klaster (Alharthi & Abdullah, 2022) Pendekatan ini dipilih karena efisiensi dan kemampuannya dalam menangani dataset berukuran besar. Proses clustering dilakukan melalui tahapan iteratif, sebagaimana diperlihatkan pada Gambar 2. Clusterisasi data



Gambar 2. Clusterisasi data

Setelah proses clustering dilakukan, evaluasi optimalitas model dilakukan menggunakan *silhouette score*, sebuah metrik yang mengukur sejauh mana data dalam satu klaster memiliki kemiripan tinggi satu sama lain, dan sejauh mana klaster tersebut berbeda dengan klaster lainnya. Berdasarkan hasil penghitungan, nilai silhouette score optimal yang diperoleh adalah 0,291, yang menunjukkan kualitas pengelompokan data pada konfigurasi terbaik. Hasil ini menghasilkan 19 klaster, dengan setiap klaster menggambarkan pola spesifik dalam dataset.

3.6 Pengujian Data (Data Testing)

Pengujian data merupakan langkah yang sesuai dengan tujuan penelitian untuk memastikan kualitas data sebelum analisis dilakukan. Proses pengujian ini mencakup pemeriksaan kesesuaian format dan tipe data guna meminimalkan kesalahan selama proses analisis. Sebagai bagian dari langkah ini, peneliti melakukan verifikasi dengan menarik laporan dari tabel data, seperti yang disajikan pada Gambar 3. Report Data Table

Demam meriang >1 bulan	DM	Berkeringat malam hari tanpa kegiatan	Sesak Nafas	Usia	Batuk	Lama Pengobatan (Hari)
1	YA	TIDAK	TIDAK	28	YA	179
2	YA	TIDAK	YA	64	YA	178
3	YA	TIDAK	YA	62	YA	172
4	TIDAK	TIDAK	YA	62	YA	153
5	TIDAK	TIDAK	TIDAK	38	YA	141
6	TIDAK	TIDAK	YA	60	YA	170
7	TIDAK	TIDAK	YA	78	YA	157
8	YA	TIDAK	YA	47	YA	178
9	YA	TIDAK	YA	54	YA	178
10	YA	TIDAK	YA	35	YA	185
11	TIDAK	TIDAK	TIDAK	49	YA	180
12	TIDAK	TIDAK	YA	38	YA	173
13	TIDAK	TIDAK	YA	32	YA	155
14	TIDAK	TIDAK	YA	22	YA	179
15	YA	TIDAK	TIDAK	15	YA	181
16	TIDAK	TIDAK	YA	55	YA	175

Gambar 3. Report Data Table

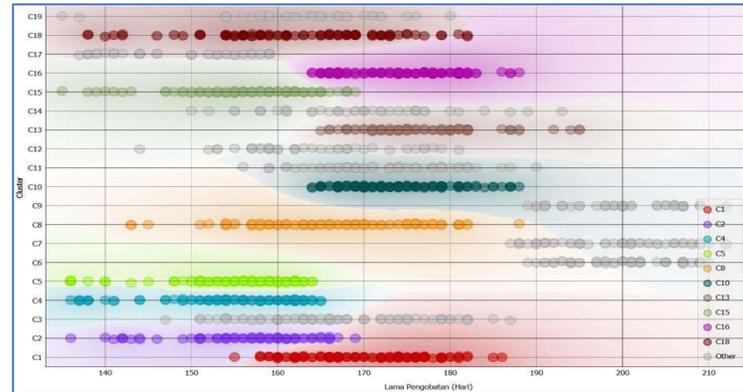
Hasil dari laporan tabel data menunjukkan bahwa tidak terdapat nilai yang hilang (missing values). Dengan demikian, dapat disimpulkan bahwa pengujian data telah berhasil, sehingga analisis data dapat dilanjutkan ke tahap berikutnya

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Setelah proses clustering dilakukan, evaluasi optimalitas model dilakukan menggunakan *silhouette* score, sebuah metrik yang mengukur sejauh mana data dalam satu kluster memiliki kemiripan tinggi satu sama lain, dan sejauh mana kluster tersebut berbeda dengan kluster lainnya. Berdasarkan hasil penghitungan, nilai *silhouette* score optimal yang diperoleh adalah 0,291, yang menunjukkan kualitas pengelompokan data pada konfigurasi terbaik. Hasil ini menghasilkan 19 kluster, dengan setiap kluster menggambarkan pola spesifik dalam dataset. Data dikelompokkan ke dalam 19 kluster untuk mengidentifikasi pola berdasarkan gejala dan durasi pengobatan.

Untuk memberikan pemahaman yang lebih jelas mengenai hasil clustering, visualisasi hasil dilakukan menggunakan Scatter Plot pada Gambar 4. Scater Plot Cluster. Plot ini menggambarkan distribusi data dalam kluster berdasarkan rentang lama pengobatan pasien TB. Visualisasi ini tidak hanya memberikan gambaran distribusi data dalam kluster tetapi juga mempermudah identifikasi pola-pola utama yang berkaitan dengan durasi pengobatan.



Gambar 4. Scater Plot Cluster

4.2 Pembahasan

Kluster C1

Kluster C1 terdiri dari 106 pasien dengan rentang usia 18 hingga 48 tahun dan durasi pengobatan antara 156 hingga 186 hari. Analisis menunjukkan bahwa gejala demam meriang lebih dari satu bulan terdeteksi pada 100% pasien, mencerminkan prevalensi yang sangat tinggi dan distribusi yang merata di seluruh kelompok usia dan durasi pengobatan. Gejala sesak napas dan batuk juga teridentifikasi pada 100% pasien, menunjukkan bahwa kedua gejala ini merupakan indikator utama dalam kluster ini. Sebaliknya, gejala diabetes mellitus (DM) hanya terdeteksi pada 2% pasien, menunjukkan distribusi yang sangat terbatas. Secara keseluruhan, 80% pasien mengalami gejala yang diamati, menandakan bahwa kluster ini memiliki karakteristik gejala yang dominan dan konsisten.

Kluster C2

Kluster C2 melibatkan 84 pasien dengan rentang usia 50 hingga 80 tahun dan durasi pengobatan antara 136 hingga 168 hari. Dalam kluster ini, gejala demam meriang lebih dari satu bulan hanya terdeteksi pada 2% pasien, menunjukkan distribusi yang sangat terbatas. Gejala diabetes mellitus ditemukan pada 1% pasien, yang juga menunjukkan prevalensi yang rendah. Meskipun gejala berkeringat malam tanpa aktivitas terdeteksi pada 4% pasien, gejala sesak napas tidak ditemukan sama sekali. Namun, gejala

batuk menunjukkan distribusi optimal, terdeteksi pada 100% pasien. Secara keseluruhan, hanya 21% dari pasien di klaster ini mengalami gejala yang terdistribusi secara merata, menandakan bahwa klaster ini memiliki karakteristik gejala yang tidak merata.

Klaster C3

Klaster C3 mencakup 84 pasien dengan rentang usia 40 hingga 90 tahun dan durasi pengobatan antara 150 hingga 180 hari. Dalam analisis ini, gejala demam meriang lebih dari satu bulan tidak terdeteksi sama sekali, menunjukkan bahwa gejala ini tidak menjadi perhatian dalam klaster ini. Gejala diabetes mellitus teridentifikasi pada 5% pasien, sementara gejala berkeringat malam tanpa aktivitas tidak ditemukan. Namun, gejala sesak napas dan batuk menunjukkan distribusi optimal, terdeteksi pada 100% pasien. Dengan total persentase sebaran gejala mencapai 41%, klaster ini menunjukkan bahwa meskipun beberapa gejala tidak terdistribusi, sesak napas dan batuk menjadi gejala yang paling dominan.

Klaster C4

Klaster C4 terdiri dari 84 pasien dengan rentang usia 48 hingga 80 tahun dan durasi pengobatan antara 136 hingga 164 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan tidak terdeteksi, sedangkan gejala diabetes mellitus ditemukan pada 3% pasien. Gejala berkeringat malam tanpa aktivitas terdeteksi pada 95% pasien, menunjukkan prevalensi yang tinggi. Gejala sesak napas juga menunjukkan distribusi yang sangat optimal, terdeteksi pada 98% pasien, sementara gejala batuk teridentifikasi pada 100% pasien. Dengan total persentase sebaran gejala mencapai 59%, klaster ini menunjukkan bahwa gejala batuk dan sesak napas adalah yang paling umum, sementara gejala lainnya memiliki distribusi yang lebih terbatas.

Klaster C5

Klaster C5 melibatkan 91 pasien dengan rentang usia 16 hingga 46 tahun dan durasi pengobatan antara 136 hingga 162 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan tidak terdeteksi, dan gejala diabetes mellitus hanya ditemukan pada 1% pasien. Namun, gejala berkeringat malam tanpa aktivitas terdistribusi secara merata, terdeteksi pada 100% pasien, dan gejala sesak napas juga menunjukkan prevalensi tinggi, terdeteksi pada 99% pasien. Gejala batuk teridentifikasi pada 98% pasien. Dengan total persentase distribusi gejala mencapai 60%, klaster ini menunjukkan bahwa meskipun beberapa gejala memiliki prevalensi rendah, gejala berkeringat malam, sesak napas, dan batuk menjadi yang paling umum.

Klaster C6

Klaster C6 terdiri dari 58 pasien dengan rentang usia 16 hingga 44 tahun dan durasi pengobatan antara 190 hingga 210 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 45% pasien, menunjukkan distribusi yang cukup merata. Gejala diabetes mellitus tidak ditemukan sama sekali, sementara gejala berkeringat malam tanpa aktivitas terdeteksi pada 72% pasien. Gejala sesak napas menunjukkan prevalensi tinggi, terdeteksi pada 93% pasien, dan gejala batuk juga terdistribusi secara merata. Dengan total persentase sebaran gejala mencapai 62%, klaster ini menunjukkan bahwa meskipun ada gejala dengan prevalensi rendah, sesak napas dan batuk menjadi gejala yang paling umum.

Klaster C7

Klaster C7 melibatkan 53 pasien dengan rentang usia 30 hingga 80 tahun dan durasi pengobatan antara 188 hingga 212 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 4% pasien, dan gejala berkeringat malam tanpa aktivitas teridentifikasi pada 17% pasien. Gejala sesak napas dan batuk juga terdeteksi pada 17% pasien. Namun, gejala

batuk menunjukkan distribusi yang lebih baik, terdeteksi pada 87% pasien. Dengan total persentase sebaran gejala mencapai 23%, klaster ini menunjukkan ketidakteraturan dalam distribusi gejala, dengan batuk sebagai gejala yang paling dominan.

Klaster C8

Klaster C8 mencakup 122 pasien dengan rentang usia 50 hingga 80 tahun dan durasi pengobatan antara 150 hingga 180 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 100% pasien, menunjukkan distribusi yang optimal. Gejala diabetes mellitus ditemukan pada 3% pasien, sedangkan gejala berkeringat malam tanpa aktivitas terdeteksi pada 100% pasien. Gejala sesak napas menunjukkan prevalensi tinggi, terdeteksi pada 94% pasien, dan gejala batuk teridentifikasi pada 98% pasien. Dengan total persentase distribusi gejala mencapai 79%, klaster ini menunjukkan bahwa sebagian besar gejala memiliki distribusi yang optimal, dengan gejala demam meriang dan berkeringat malam menjadi yang paling umum.

Klaster C9

Klaster C9 terdiri dari 43 pasien dengan rentang usia 50 hingga 80 tahun dan durasi pengobatan antara 150 hingga 212 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 44% pasien, menunjukkan distribusi yang cukup merata. Gejala diabetes mellitus ditemukan pada 12% pasien, sedangkan gejala berkeringat malam tanpa aktivitas terdeteksi pada 93% pasien. Gejala sesak napas dan batuk menunjukkan prevalensi tinggi, terdeteksi pada 95% pasien. Dengan total persentase sebaran gejala mencapai 68%, klaster ini menunjukkan bahwa meskipun ada gejala dengan prevalensi rendah, sesak napas, batuk, dan berkeringat malam menjadi gejala yang paling umum.

Klaster C10

Klaster C10 melibatkan 122 pasien dengan rentang usia 16 hingga 42 tahun dan durasi pengobatan antara 164 hingga 188 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan tidak terdeteksi, dan gejala diabetes mellitus hanya ditemukan pada 1% pasien. Namun, gejala berkeringat malam tanpa aktivitas terdeteksi pada 100% pasien, menunjukkan distribusi yang optimal. Gejala sesak napas ditemukan pada 95% pasien, dan gejala batuk teridentifikasi pada 100% pasien. Dengan total persentase distribusi gejala mencapai 59%, klaster ini menunjukkan bahwa meskipun beberapa gejala memiliki prevalensi rendah, gejala berkeringat malam, sesak napas, dan batuk menjadi yang paling umum.

Klaster C11

Klaster C11 terdiri dari 70 pasien dengan rentang usia 52 hingga 88 tahun dan durasi pengobatan antara 156 hingga 190 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 4% pasien, dan gejala diabetes mellitus ditemukan pada 9% pasien. Gejala berkeringat malam tanpa aktivitas tercatat pada 3% pasien, sedangkan gejala sesak napas hanya terdeteksi pada 1% pasien. Namun, gejala batuk menunjukkan distribusi yang sangat baik, terdeteksi pada 99% pasien. Dengan total persentase sebaran gejala mencapai 23%, klaster ini menunjukkan bahwa hanya gejala batuk yang memiliki distribusi optimal, sementara gejala lainnya menunjukkan prevalensi yang sangat terbatas.

Klaster C12

Klaster C12 melibatkan 40 pasien dengan rentang usia 40 hingga 80 tahun dan durasi pengobatan antara 144 hingga 180 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 100% pasien, menunjukkan distribusi yang optimal. Gejala diabetes mellitus hanya ditemukan pada 3% pasien, dan gejala berkeringat malam tanpa aktivitas tidak terdeteksi. Gejala sesak napas

terdeteksi pada 90% pasien, dan gejala batuk teridentifikasi pada 100% pasien. Dengan total persentase sebaran gejala mencapai 59%, klaster ini menunjukkan bahwa gejala batuk, sesak napas, dan demam meriang memiliki distribusi yang optimal, sementara gejala DM dan berkeringat malam tidak menunjukkan prevalensi yang signifikan.

Klaster C13

Klaster C13 terdiri dari 91 pasien dengan rentang usia 16 hingga 50 tahun dan durasi pengobatan antara 166 hingga 194 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 3% pasien, dan gejala diabetes mellitus ditemukan pada 2% pasien. Gejala berkeringat malam tanpa aktivitas tidak ditemukan, sedangkan gejala sesak napas terdeteksi pada 1% pasien. Namun, gejala batuk menunjukkan distribusi yang sangat baik, terdeteksi pada 97% pasien. Dengan total persentase sebaran gejala mencapai 21%, klaster ini menunjukkan bahwa meskipun gejala batuk memiliki prevalensi yang tinggi, gejala lainnya menunjukkan distribusi yang sangat terbatas.

Klaster C14

Klaster C14 melibatkan 42 pasien dengan rentang usia 20 hingga 60 tahun dan durasi pengobatan antara 150 hingga 190 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 90% pasien, menunjukkan distribusi yang merata. Gejala diabetes mellitus ditemukan pada 12% pasien, sedangkan gejala berkeringat malam tanpa aktivitas terdeteksi pada 90% pasien. Gejala sesak napas tidak ditemukan sama sekali, sementara gejala batuk menunjukkan distribusi yang sangat baik, terdeteksi pada 100% pasien. Dengan total persentase sebaran gejala mencapai 59%, klaster ini menunjukkan bahwa gejala batuk, demam meriang, dan berkeringat malam memiliki distribusi yang optimal, sementara gejala DM dan sesak napas tidak terdeteksi.

Klaster C15

Klaster C15 terdiri dari 87 pasien dengan rentang usia 16 hingga 42 tahun dan durasi pengobatan antara 136 hingga 168 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 10% pasien, dan gejala diabetes mellitus ditemukan pada 6% pasien. Gejala berkeringat malam tanpa aktivitas hanya terdeteksi pada 2% pasien, dan gejala sesak napas tidak ditemukan. Namun, gejala batuk menunjukkan distribusi yang sangat baik, terdeteksi pada 97% pasien. Dengan total persentase sebaran gejala mencapai 23%, klaster ini menunjukkan bahwa meskipun gejala tidak terdistribusi secara merata, gejala batuk menjadi yang paling dominan.

Klaster C16

Klaster C16 melibatkan 129 pasien dengan rentang usia 50 hingga 80 tahun dan durasi pengobatan antara 164 hingga 188 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan tidak terdeteksi, dan gejala diabetes mellitus hanya ditemukan pada 1% pasien. Namun, gejala berkeringat malam tanpa aktivitas terdeteksi pada 100% pasien, menunjukkan distribusi yang optimal. Gejala sesak napas dan batuk juga menunjukkan prevalensi tinggi, terdeteksi pada 99% pasien. Dengan total persentase sebaran gejala mencapai 60%, klaster ini menunjukkan bahwa gejala berkeringat malam, sesak napas, dan batuk memiliki distribusi yang merata, sementara gejala demam meriang dan diabetes mellitus menunjukkan prevalensi yang sangat rendah.

Klaster C17

Klaster C17 terdiri dari 41 pasien dengan rentang usia 20 hingga 50 tahun dan durasi pengobatan antara 138 hingga 158 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 100% pasien, menunjukkan distribusi yang merata. Gejala diabetes mellitus hanya ditemukan pada 2%

pasien, sedangkan gejala berkeringat malam tanpa aktivitas terdeteksi pada 90% pasien. Gejala sesak napas dan batuk juga terdeteksi pada 100% pasien, menunjukkan bahwa kedua gejala ini merupakan indikator utama dalam klaster ini. Dengan total persentase sebaran gejala yang cukup tinggi, klaster ini menunjukkan bahwa meskipun gejala DM jarang muncul, gejala lainnya memiliki prevalensi yang baik.

Klaster C18

Klaster C18 melibatkan 89 pasien dengan rentang usia 16 hingga 46 tahun dan durasi pengobatan antara 140 hingga 180 hari. Dalam klaster ini, gejala demam meriang lebih dari satu bulan terdeteksi pada 13% pasien, sedangkan gejala diabetes mellitus tidak ditemukan. Gejala berkeringat malam tanpa aktivitas juga tidak terdeteksi, sementara gejala sesak napas menunjukkan prevalensi tinggi, terdeteksi pada 100% pasien. Gejala batuk juga teridentifikasi pada 99% pasien. Dengan total persentase sebaran gejala mencapai 42%, klaster ini menunjukkan bahwa meskipun ada gejala dengan prevalensi rendah, sesak napas dan batuk menjadi gejala yang paling umum.

Klaster C19

Klaster C19 terdiri dari 37 pasien dengan rentang usia 30 hingga 70 tahun dan durasi pengobatan antara 140 hingga 180 hari. Dalam klaster ini, tidak ada gejala demam meriang, diabetes mellitus, berkeringat malam tanpa aktivitas, sesak napas, atau batuk yang terdeteksi. Dengan total persentase sebaran gejala mencapai 0%, klaster ini menunjukkan bahwa pasien dalam klaster ini tidak mengalami gejala yang dapat dihubungkan dengan kondisi mereka. Hal ini menandakan bahwa klaster C19 memiliki karakteristik unik di mana tidak ada gejala yang teridentifikasi, meskipun pasien berada dalam rentang usia dan durasi pengobatan yang bervariasi.

5. CONCLUSION

Analisis distribusi gejala tuberkulosis (TB) di antara 19 klaster pasien menunjukkan variasi yang signifikan dalam prevalensi dan pola gejala berdasarkan usia dan durasi pengobatan. Setiap klaster memiliki karakteristik unik yang mencerminkan tantangan dan kebutuhan spesifik dalam penanganan TB.

1. **Klaster C1** menunjukkan prevalensi tinggi untuk gejala demam meriang, sesak napas, dan batuk, dengan 100% pasien mengalami gejala tersebut. Hal ini menandakan bahwa pasien dalam kelompok usia produktif ini mungkin memerlukan perhatian medis yang lebih intensif untuk mengelola gejala yang parah.
2. **Klaster C2 dan C3** memperlihatkan distribusi gejala yang lebih terbatas, dengan gejala batuk yang mendominasi di C2, sementara C3 menunjukkan prevalensi tinggi untuk sesak napas dan batuk. Ini menunjukkan bahwa meskipun jumlah pasien di kedua klaster ini relatif sama, respons terhadap pengobatan dan manifestasi gejala dapat bervariasi secara signifikan.
3. **Klaster C4 dan C5** menunjukkan distribusi gejala yang lebih optimal, terutama untuk batuk dan sesak napas, dengan C4 memiliki 100% pasien yang mengalami batuk. Hal ini menunjukkan bahwa pasien di klaster ini mungkin lebih responsif terhadap pengobatan, tetapi tetap memerlukan pemantauan untuk gejala lain yang mungkin muncul.
4. **Klaster C6 dan C7** menunjukkan prevalensi tinggi untuk sesak napas dan batuk, tetapi dengan variasi dalam gejala lain seperti demam meriang dan diabetes mellitus yang tidak terdeteksi secara signifikan. Ini menunjukkan bahwa

meskipun ada gejala dominan, ada kemungkinan adanya gejala lain yang tidak teridentifikasi yang dapat mempengaruhi kualitas hidup pasien.

5. **Klaster C8 dan C9** menunjukkan distribusi gejala yang lebih merata, dengan batuk dan sesak napas sebagai gejala utama. Klaster C8, dengan jumlah pasien yang lebih besar, menunjukkan bahwa pengelolaan gejala harus dilakukan secara komprehensif untuk meningkatkan kualitas hidup pasien.
6. **Klaster C10 hingga C19** menunjukkan variasi yang lebih besar dalam distribusi gejala, dengan beberapa klaster seperti C19 tidak menunjukkan gejala signifikan sama sekali. Hal ini menandakan bahwa ada kemungkinan pasien dalam klaster ini mungkin tidak terdiagnosis dengan baik atau memiliki kondisi kesehatan yang berbeda yang mempengaruhi gejala TB.

Dari hasil analisis ini menekankan pentingnya pendekatan yang disesuaikan dalam penanganan TB, dengan mempertimbangkan faktor usia, durasi pengobatan, dan pola gejala yang muncul. Penelitian lebih lanjut diperlukan untuk memahami interaksi antara faktor-faktor ini dan untuk merumuskan strategi intervensi yang lebih efektif. Dengan demikian, upaya penanggulangan TB dapat lebih terarah dan berbasis bukti, yang pada akhirnya akan meningkatkan hasil kesehatan bagi pasien di seluruh klaster.

REFERENCES

Abbas, M. A., Munir, K., Raza, A., Samee, N. A.,

Jamjoom, M. M., & Ullah, Z. (2024). Novel Transformer Based Contextualized Embedding and Probabilistic Features for Depression Detection From Social Media. In *IEEE Access*

(Vol. 12, pp. 54087–54100). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2024.3387695>

Acheampong, F. A. (2020). Text-based emotion detection: Advances, challenges, and opportunities. In *Engineering Reports* (Vol. 2, Issue 7). <https://doi.org/10.1002/eng2.12189>

Ahmadian, S., Bateni, M., Esfandiari, H., Lattanzi, S., Monemizadeh, M., & Norouzi-Fard, A. (2024). Resilient k-Clustering. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 29–38). ACM. <https://doi.org/10.1145/3637528.3671888>

Alamoodi, A. H. (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. In *Expert Systems with Applications* (Vol. 167). <https://doi.org/10.1016/j.eswa.2020.114155>

Alharthi, M. M., & Abdullah, M. A. (2022). Mining Data Streams using Clustering Techniques. In *International Journal of Computer Applications* (Vol. 184, Issue 7, pp. 9–15). Foundation of Computer Science. <https://doi.org/10.5120/ijca2022922027>

Alim, D. A. (2024). Review of: “Application of Data Mining Combined with K-means Clustering Algorithm in Enterprises’ Risk Audit.” Qeios Ltd. <https://doi.org/10.32388/t7cxyw>

- Alkattan, H., Abbas, N. R., Adelaja, O. A., Abotaleb, M., & Ali, G. (2024). Data Mining Utilizing Various Leveled Clustering Procedures on the Position of Workers in a Data Innovation Firm. In *Mesopotamian Journal of Computer Science* (Vol. 2024, pp. 122–127). Mesopotamian Academic Press.
<https://doi.org/10.58496/mjcs/2024/008>
- Annas, M., & Wahab, S. N. (2023). Data Mining Methods: K-Means Clustering Algorithms. In *International Journal of Cyber and IT Service Management* (Vol. 3, Issue 1, pp. 40–47). Pandawan.
<https://doi.org/10.34306/ijcitsm.v3i1.122>
- Banait, S. S., & SANE, D. S. S. (2022). An Efficient Clustering Technique for Big Data Mining. In *International Journal of Next-Generation Computing*. Perpetual Innovation Media Pvt. Ltd. <https://doi.org/10.47164/ijngc.v13i3.842>
- Biswas, N. (2024). *Review of: "Application of Data Mining Combined with K-means Clustering Algorithm in Enterprises' Risk Audit."* Qeios Ltd. <https://doi.org/10.32388/e03be8>
- Global tuberculosis report 2022*. (n.d.). Retrieved January 27, 2024, from
- <https://www.who.int/publications/i/item/9789240061729>
- Harpad, B., Fahmi, M., Pahrudin, P., & Andrea, R. (2024). Pengelompokan Siswa Layak Penerima Beasiswa dengan Menerapkan Algoritma K-Means Clustering Data Mining. In *JURNAL MEDIA INFORMATIKA BUDIDARMA* (Vol. 8, Issue 2, pp. 913–913). STMIK Budi Darma.
<https://doi.org/10.30865/mib.v8i2.7394>
- Hasheminejad, S. M. H., & Khorrami, M. (2021). Clustering of bank customers based on lifetime value using data mining methods. In *Intelligent Decision Technologies* (Vol. 14, Issue 4, pp. 507–515). SAGE Publications.
<https://doi.org/10.3233/idt-190176>
- Li, M., Frank, E., & Pfahringer, B. (2022). Large scale K-means clustering using GPUs. In *Data Mining and Knowledge Discovery* (Vol. 37, Issue 1, pp. 67–109). Springer Science and Business Media LLC.
<https://doi.org/10.1007/s10618-022-00869-6>
- SITK Penabulu-STPI*. (n.d.). Retrieved May 27, 2024, from <http://sitk.online/>