

ANALISIS KLASSTER PASIEN RUMAH SAKIT JiWA KOTA KENDARI TAHUN 2022-2023 IDENTIFIKASI POLA DAN KARAKTERISTIK MENGUNAKAN METODE DBSCAN

Muh Saleh Badawi, Fellonnisa Suria, Elfiani, Widya Dwi Tsunami, Rizal Adi Saputra

Teknik Informatika, Universitas Halu Oleo Kendari,

Kampus Hijau Bumi Tridharma, Anduonohu, Kec. Kambu, Kota Kendari, Sulawesi Tenggara
kemalbadawi@gmail.com

ABSTRAK

Rumah Sakit Jiwa Kota Kendari berperan penting dalam menangani pasien dengan masalah kesehatan mental di Kota Kendari. Penelitian ini bertujuan untuk mengidentifikasi pola dan karakteristik pasien selama tahun 2022 dan 2023. Peningkatan jumlah pasien yang signifikan dalam dua tahun terakhir menimbulkan tantangan dalam pengelolaan dan perencanaan sumber daya kesehatan mental. Tanpa pemahaman yang jelas tentang karakteristik dan pola pasien, sulit untuk merancang intervensi yang efisien. Penelitian ini menggunakan metode DBSCAN (Density-Based Spatial Clustering of Applications with Noise) untuk menganalisis data pasien berdasarkan variabel usia, jenis kelamin, status pasien (baru/lama), dan bulan kunjungan. Hasil pengujian menunjukkan enam kluster utama, termasuk satu kluster outlier, dengan rata-rata Silhouette Score sebesar 0.365 yang mengindikasikan kualitas klusterisasi sedang. Setiap kluster menunjukkan perbedaan karakteristik, seperti dominasi pasien baru laki-laki pada kluster tertentu, pasien lama perempuan pada kluster lainnya, dan outlier dengan rentang usia bervariasi yang membutuhkan perhatian khusus. Temuan ini memberikan pemahaman mendalam tentang distribusi demografis dan pola kunjungan pasien. Hasil penelitian dapat digunakan untuk merancang strategi pelayanan kesehatan jiwa yang lebih efektif, mengalokasikan sumber daya secara efisien, dan mengembangkan program perawatan yang lebih terarah di Rumah Sakit Jiwa Kota Kendari.

Kata kunci : Rumah Sakit Jiwa Kota Kendari, Masalah mental, Pola dan karakteristik pasien, DBSCAN, Kluster pasien

1. PENDAHULUAN

Rumah Sakit Jiwa Kota Kendari memiliki peran penting dalam memberikan perawatan bagi individu yang mengalami gangguan kesehatan mental. Informasi mengenai pasien yang diterima di rumah sakit ini adalah sumber data berharga yang dapat membantu memahami pola dan karakteristik pasien. Analisis data ini diperlukan untuk mendukung perencanaan dan peningkatan layanan kesehatan pada rumah sakit jiwa kota kendari.

Penelitian ini bertujuan dapat mengidentifikasi pola dan karakteristik kluster pasien berdasarkan data yang ada. Dengan mengidentifikasi kluster-kluster ini, kami berharap dapat menemukan tren umum dalam data pasien yang dapat digunakan untuk meningkatkan kualitas layanan perawatan di masa mendatang. Dalam upaya meningkatkan kualitas pelayanan kesehatan jiwa, pemahaman mendalam tentang karakteristik dan pola pasien menjadi krusial. Analisis terhadap data pasien dapat mengungkapkan informasi berharga yang selama ini mungkin terlewatkan. Metode analisis kluster, khususnya DBSCAN (Density-Based Spatial Clustering of Applications with Noise), menawarkan pendekatan yang menjanjikan untuk mengidentifikasi pola-pola tersembunyi dalam data pasien.

Di antara berbagai jenis algoritme *clustering*, *density based clustering* ini dapat menentukan cluster dari bentuk data yang tidak beraturan dan dapat menangani noise secara efektif dan DBSCAN memiliki kelebihan, diantaranya metode DBSCAN ini

secara signifikan lebih efektif dalam menemukan cluste dengan bentuk yang berubah-ubah, dan juga dapat menemukan kluster yang memiliki bentuk yang tidak tentu[1]. Metode ini belum banyak diterapkan dalam konteks analisis data kesehatan jiwa di Indonesia, sehingga penerapannya dalam studi ini diharapkan dapat memberikan perspektif baru dalam memahami karakteristik pasien rumah sakit jiwa.

Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang kesehatan mental, khususnya dalam perencanaan dan pelaksanaan layanan kesehatan mental di Rumah Sakit Jiwa Kota Kendari. Dengan memahami pola dan karakteristik dari kluster pasien, pihak rumah sakit dapat merancang strategi perawatan yang lebih tepat sasaran dan efektif, sehingga dapat meningkatkan kualitas hidup pasien dan masyarakat secara keseluruhan.

2. TINJAUAN PUSTAKA

Analisis cluster merupakan salah satu alat yang krusial dalam pengolahan data statistik untuk melakukan analisis. Metode ini secara otomatis mengelompokkan objek ke dalam kelompok berdasarkan kesamaan yang ada di antara mereka. [2]. Dalam konteks kesehatan, analisis kluster dapat membantu mengidentifikasi pola-pola tertentu dalam data pasien, yang dapat digunakan untuk meningkatkan kualitas pelayanan perawatan dan pengelolaan pasien.

Salah satu metode clustering yang dapat digunakan adalah DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Algoritma DBSCAN termasuk dalam kategori algoritma nonparametrik dalam pembelajaran tanpa pengawasan (unsupervised learning), sehingga tidak memerlukan asumsi khusus selama proses analisis. Secara konseptual, DBSCAN dapat membentuk cluster dengan bentuk yang fleksibel dan tidak terikat pada pola tertentu, sehingga tidak harus berbentuk bulat. Selain itu, algoritma ini efektif dalam menangani cluster yang mengandung noise atau pencilan. DBSCAN bekerja dengan cara memperluas area berkepadatan tinggi menjadi cluster menggunakan dua parameter utama yang harus ditentukan dengan tepat, yaitu radius jangkauan (ϵ) dan jumlah minimum objek (MinObj). Kedua parameter ini digunakan untuk menentukan apakah suatu area memiliki kepadatan yang cukup untuk dianggap sebagai sebuah cluster [3]. Metode DBSCAN memiliki kemampuan untuk mengidentifikasi klaster dengan bentuk yang tidak teratur dan dapat menangani data yang mengandung noise, menjadikannya alat yang efektif untuk analisis data yang kompleks seperti data pasien rumah sakit jiwa.

2.1. Clustering

Clustering adalah teknik pengelompokan record dalam basis data berdasarkan kriteria tertentu. Hasil dari proses clustering disajikan kepada pengguna akhir untuk memberikan pemahaman mengenai pola atau struktur yang terdapat dalam basis data. Konsep dasar clustering adalah mengelompokkan sejumlah objek ke dalam cluster, di mana cluster yang ideal memiliki tingkat kesamaan yang tinggi antar objek di dalam cluster tersebut, serta tingkat perbedaan yang signifikan dengan objek di cluster lainnya. Metode clustering hadir dikarenakan kondisi berlimpahnya data yang merupakan akumulasi data transaksi yang terekam selama bertahun-tahun[4].

Clustering merupakan proses pengelompokan data yang mirip ke dalam kelompok-kelompok berbeda. Lebih spesifik, clustering membagi sebuah dataset menjadi beberapa subset sehingga setiap subset memiliki fungsi atau nilai tertentu. Proses ini juga dapat diartikan sebagai cara untuk mengelompokkan

dan memisahkan data menjadi beberapa bagian hingga terbentuk struktur yang terorganisir.[5].

2.2. Metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN adalah algoritma clustering berbasis kepadatan (density-based) yang bekerja dengan memperluas area dengan kepadatan tinggi untuk membentuk cluster. Algoritma ini juga mampu menangani cluster dengan bentuk tidak teratur pada database spasial yang mengandung noise. Dalam metode ini, cluster didefinisikan sebagai himpunan maksimal dari titik-titik yang saling terhubung berdasarkan kepadatan (density-connected). DBSCAN menggunakan dua parameter utama, yaitu Eps (radius maksimum dari neighborhood) dan MinPts (jumlah minimum titik dalam Eps-neighborhood suatu titik).[6]. DBSCAN memiliki keunggulan dalam proses clustering yaitu DBSCAN mengelompokkan berdasar kepada kepadatan data atau density sehingga unggul dalam pengelompokan data yang besar [7].

2.3. Analisis Pola dan Karakteristik dalam Data Kesehatan

Analisis pola dan karakteristik dalam data kesehatan adalah metode krusial yang memungkinkan dokter dan perawat memperoleh wawasan mengenai bagaimana penyakit berkembang, bagaimana pasien berperilaku, dan seberapa efektif tindakan medis yang dilakukan pada waktu yang akan datang. Dengan menerapkan metode pengolahan informasi, berbagai pola penting dan keunikan dalam kelompok pasien dapat ditemukan. Temuan-temuan ini selanjutnya bisa dimanfaatkan untuk memperbaiki mutu layanan kesehatan di rumah sakit.

3. METODE PENELITIAN

3.1. Data

Data yang di gunakan dalam penelitian ini yaitu data pasien rumah sakit jiwa kota kendari tahun 2022-2023, yang dimana data di proleh dari rumah sakit jiwa kota kendari, terdapat 18280 pasien dalam dataset, yang dimana data tersebut meliputi beberapa fitur penting diantaranya usia, jenis kelamin, bulan kunjungan, dan status pasien. Adapun data yang digunakan ditampilkan pada Tabel 1 berikut:

Tabel 1 Data Pasien Rumah Sakit Jiwa Kota Kendari

No	Pasien Lama	Pasien Baru	Bulan	Usia	Jenis Kelamin
1	0	1	1	56	0
2	0	1	1	51	0
3	0	1	1	35	1
...	...	...	...	...	...
18379	1	0	12	50	0
18380	0	1	12	27	1

3.2. Algoritma DBSCAN

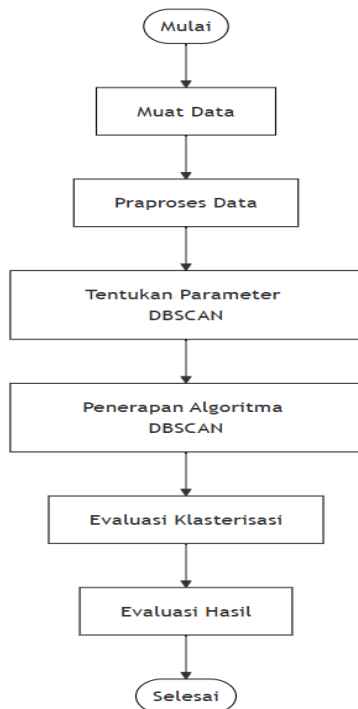
Penelitian ini menggunakan metode clustering DBSCAN untuk mengelompokkan pasien di Rumah Sakit Jiwa Kota Kendari. Algoritma DBSCAN adalah salah satu pendekatan clustering yang berbasis pada

kepadatan. Sebagai metode deteksi berbasis klaster yang terkenal, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) mampu mengidentifikasi kelompok data dengan mempertimbangkan kepadatan. Algoritma ini

merupakan salah satu metode utama dalam pengelompokan data dalam proses penambangan data. [8]. DBSCAN memiliki kemampuan untuk menemukan suatu kelompok ukuran dan bentuk variabel dan juga akan mendeteksi noise, dari sejumlah data besar yang mengandung noise dan outlier.

3.3. Metode Analisis Data

Langkah-langkah analisis data pada penelitian ini akan dijelaskan secara sistematis, yang mana akan ditampilkan dalam bentuk flowchart untuk mempermudah pemahaman alur proses analisis:



Gambar 1. Flowchart Proses Analisis Clustering Data Menggunakan Algoritma DBSCAN

a. Praproses Data

Tahap praproses dilakukan sebagai tahap awal dan tahap penting dalam penelitian [9]. Langkah pertama dalam praproses data adalah mengisi missing values. Missing values diisi menggunakan metode forward fill untuk memastikan tidak adanya data yang hilang yang dapat mempengaruhi hasil analisis. Selanjutnya, fitur yang relevan untuk analisis kluster dipilih, termasuk usia, jenis kelamin, bulan kunjungan, dan status pasien (lama atau baru). Fitur-fitur ini kemudian distandarisasi menggunakan StandardScaler untuk memastikan semua fitur berada pada skala yang sama, sehingga tidak ada fitur yang mendominasi yang lain.

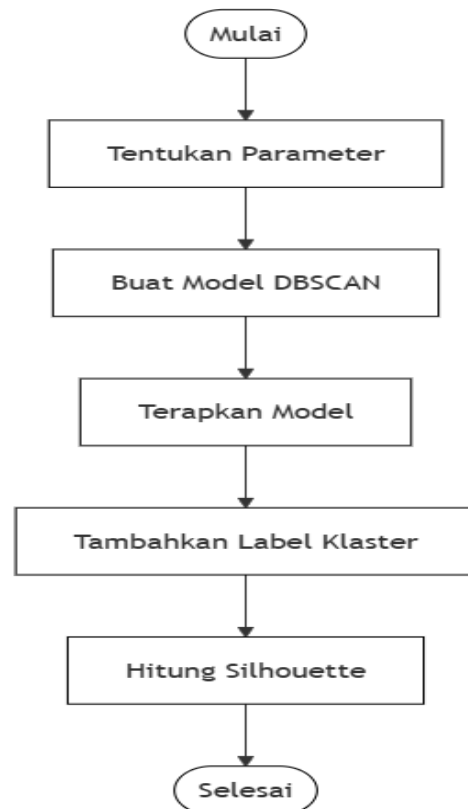
b. Parameter DBSCAN

Setelah praproses data, langkah berikutnya adalah menentukan parameter utama dalam DBSCAN, yaitu epsilon (eps) dan min_samples. Epsilon menentukan radius maksimal dari lingkungan suatu titik untuk dimasukkan ke dalam kluster. Dalam penelitian ini, epsilon diset pada 0.5.

Min_samples menentukan jumlah minimum titik yang diperlukan untuk membentuk suatu kluster, dan dalam penelitian ini, min_samples diset pada Pemilihan parameter ini penting karena sangat mempengaruhi hasil klusterisasi.

c. Penerapan Algoritma DBSCAN

Setelah parameter ditentukan, model DBSCAN diterapkan pada data yang telah distandarisasi untuk mengidentifikasi kluster dalam data pasien[10]. Algoritma ini akan mengelompokkan titik-titik data ke dalam kluster berdasarkan kerapatan lokal data. Hasil klusterisasi berupa label kluster kemudian ditambahkan ke data asli, memungkinkan identifikasi setiap titik data dengan kluster tertentu atau sebagai noise jika tidak termasuk dalam kluster manapun. Alur dari proses ini akan ditampilkan dalam bentuk flowchart untuk mempermudah pemahaman langkah-langkah yang dilakukan dalam analisis data.



Gambar 2. Flowchart Implementasi Algoritma DBSCAN

Penjelasan detail setiap tahap DBSCAN:

a. Mulai

Tahap awal memulai proses clustering yang menandai dimulainya analisis pengelompokan data. Pada tahap ini, data sudah harus dalam kondisi bersih dan siap untuk diproses menggunakan algoritma DBSCAN.

b. Tentukan Parameter

Parameter utama DBSCAN meliputi epsilon dan min_samples yang sangat krusial dalam menentukan hasil clustering. Epsilon menentukan

radius maksimum tetangga, sementara min_samples menentukan jumlah minimum titik untuk membentuk cluster. Pemilihan parameter yang tepat akan menghasilkan cluster yang bermakna.

c. Buat Model DBSCAN

Proses ini melibatkan inisialisasi algoritma DBSCAN dengan parameter yang telah ditentukan. Model dibuat menggunakan library sklearn dengan mengimplementasikan parameter epsilon dan min_samples yang telah ditetapkan sebelumnya.

d. Terapkan Mode

Tahap implementasi dimana model DBSCAN dijalankan pada dataset. Algoritma akan mengidentifikasi core points (titik inti), border points (titik batas), dan noise points (titik noise) berdasarkan parameter yang telah ditentukan. Proses ini menghasilkan pembentukan cluster berdasarkan densitas data.

e. Tambahkan Label Kluster

Setelah model diterapkan, setiap titik data akan diberi label numerik yang mengidentifikasi keanggotaan clusternya. Noise points diberi label -1, sedangkan cluster-cluster yang terbentuk diberi label mulai dari 0. Label ini memudahkan identifikasi dan analisis hasil clustering.

f. Hitung Silhouette

Evaluasi kualitas clustering menggunakan Silhouette score yang mengukur seberapa baik setiap objek berada dalam clusternya. Nilai score berkisar antara -1 hingga 1, dimana nilai yang mendekati 1 menunjukkan clustering yang optimal. Perhitungan ini membantu validasi hasil pengelompokan.

g. Evaluasi Klasterisasi

Untuk mengevaluasi hasil klasterisasi, digunakan Silhouette Score, yang mengukur seberapa baik kluster yang terbentuk. Silhouette Score berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan kluster yang lebih baik. Evaluasi ini membantu dalam menilai kualitas kluster yang dihasilkan oleh algoritma DBSCAN.

Evaluasi klasterisasi menggunakan Silhouette Score, yang dirumuskan sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

Keterangan :

$s(i)$ = Silhouette Score untuk titik data i

$a(i)$ = jarak rata-rata antara i dan semua titik lain dalam kluster yang sama.

$b(i)$ = jarak rata-rata antara i dan semua titik dalam kluster terdekat yang berbeda.

Silhouette Score rata-rata dihitung sebagai:

$$\text{Silhouette Score} = \frac{1}{n} \sum_i s(i) \quad (2)$$

Di mana n adalah jumlah total titik data.

d. Visualisasi Hasil Klasterisasi

Berbagai visualisasi dilakukan untuk menggambarkan hasil klasterisasi. Ini termasuk scatter plot, pair plot, box plot, dan count plot untuk fitur-fitur yang relevan. Visualisasi ini membantu dalam memahami distribusi dan karakteristik kluster yang terbentuk, serta hubungan antar fitur dalam setiap kluster. Scatter plot digunakan untuk melihat distribusi titik data berdasarkan dua fitur tertentu, sedangkan pair plot memberikan gambaran hubungan antar semua fitur.

e. Analisis Karakteristik Tiap Kluster

Terakhir, dilakukan analisis deskriptif untuk setiap kluster guna mengidentifikasi karakteristik khas dari masing-masing kluster. Analisis ini meliputi statistik dasar seperti mean, median, dan standard deviation untuk setiap fitur dalam kluster. Dengan memahami karakteristik ini, dapat diidentifikasi pola-pola yang signifikan dalam data pasien, seperti perbedaan demografis atau perilaku antara kluster yang berbeda. Analisis ini sangat berguna bagi pengambilan keputusan dalam manajemen dan perawatan pasien di Rumah Sakit Jiwa Kota Kendari.

Analisis karakteristik menggunakan statistik deskriptif seperti:

Mean (Rata-rata):

$$\text{Mean} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

Median: Nilai tengah dari data yang diurutkan.

Standard Deviation (Standar Deviasi):

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (4)$$

Statistik ini digunakan untuk menggambarkan karakteristik umum dari setiap kluster.

Dengan menggunakan rumus dan metode di atas, analisis klasterisasi DBSCAN diharapkan dapat memberikan wawasan mendalam tentang pola dan karakteristik pasien Rumah Sakit Jiwa Kota Kendari.

4. HASIL DAN PEMBAHASAN

4.1. Proses Klasterisasi

Pada penelitian ini, proses klasterisasi dilakukan dengan menggunakan metode DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Berikut adalah langkah-langkah utama dalam proses klasterisasi:

a. Pengumpulan Data

Dataset yang digunakan berasal dari data pasien Rumah Sakit Jiwa Kota Kendari tahun 2022-2023. Data ini mencakup berbagai atribut seperti jumlah pasien lama dan baru, jenis kelamin, bulan masuk, serta usia pasien.

b. Praproses Data

Sebelum melakukan klusterisasi dengan DBSCAN, langkah-langkah praproses data sangat penting untuk memastikan bahwa data dalam kondisi optimal untuk analisis. Berikut langkah-langkah praproses yang dilakukan :

- Mengisi nilai yang hilang menggunakan metode forward fill.
- Memilih fitur yang relevan: Pasien Lama, Pasien Baru, Jenis Kelamin, Bulan, dan Usia.
- Melakukan standarisasi data menggunakan StandardScaler untuk menyamakan skala pada setiap fitur.

c. Menentukan Parameter DBSCAN

DBSCAN memiliki dua parameter utama yang perlu ditentukan dengan baik. Epsilon (ϵ) Jarak maksimum antar titik yang dianggap sebagai tetangga dan Min_samples yaitu Jumlah minimum titik yang diperlukan untuk membentuk sebuah kluster. Pada penelitian ini, model DBSCAN diterapkan untuk melakukan klusterisasi data pasien dengan parameter utama epsilon (eps) dan jumlah minimum sampel (min_samples). Parameter epsilon ditentukan sebesar 2.0, yang menunjukkan jarak maksimum antar titik data untuk dianggap sebagai tetangga dalam kluster yang sama. Nilai epsilon ini disesuaikan dengan karakteristik data untuk memastikan pembentukan kluster yang optimal. Selain itu, parameter min_samples ditetapkan sebesar 5, yang berarti setiap kluster harus terdiri dari minimal lima titik data untuk dianggap valid

d. Melakukan Klusterisasi

Setelah menentukan parameter epsilon (ϵ) sebesar 2.0 dan min_samples sebesar 5, proses klusterisasi dilakukan menggunakan algoritma DBSCAN. Klusterisasi bertujuan untuk mengelompokkan data pasien berdasarkan karakteristik tertentu yang dimiliki. Algoritma ini bekerja dengan prinsip kepadatan, yaitu mengidentifikasi core point, border point, dan noise dalam data. Langkah-Langkah Klusterisasi

- Identifikasi Core dan Border Point Setiap data dihitung jaraknya menggunakan metrik tertentu (misalnya, Euclidean Distance). Titik dengan jumlah tetangga dalam radius $\epsilon \geq \text{min_samples}$ dianggap core point, sedangkan titik dalam radius core point disebut border point.
- Identifikasi Noise Data yang bukan core atau border point dianggap noise dan dimasukkan ke kluster -1.
- Penerapan Algoritma DBSCAN, Implementasi algoritma dilakukan menggunakan library Python, seperti sklearn.cluster.DBSCAN.

e. Visualisasi Kluster

Untuk memahami hasil klusterisasi, dilakukan berbagai visualisasi:

- Scatter plot Usia vs Bulan dengan kluster
- Scatter plot Usia vs Jenis Kelamin
- Pairplot hubungan antar fitur dengan kluster

f. Distribusi jumlah pasien dalam setiap kluster menggunakan bar plot

- Box plot distribusi usia per kluster
- Distribusi jenis kelamin, pasien lama, dan bulan masuk per kluster

g. Evaluasi Klusterisasi

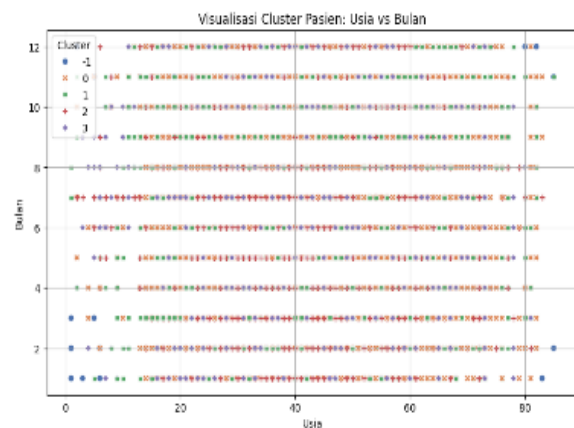
Untuk mengevaluasi kualitas kluster yang terbentuk, digunakan Silhouette Score, yang mengukur seberapa baik titik dalam suatu kluster terkelompok dengan benar.

4.2. Visualisasi Klusterisasi

Visualisasi data adalah teknik untuk mengeksplorasi dan menyajikan data dengan mengubahnya ke dalam format visual, seperti tabel atau grafik. Metode ini mempermudah analisis dan pelaporan data, serta membantu menyampaikan informasi dan bukti secara efektif, sehingga data yang kompleks menjadi lebih mudah dipahami. Beberapa visualisasi digunakan untuk memahami distribusi data dalam setiap kluster. Berikut adalah beberapa plot yang didapatkan

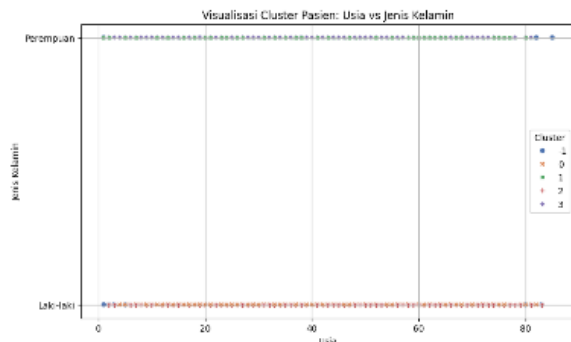
a. Scatter Plot Usia vs Bulan dengan Kluster

Menunjukkan distribusi usia terhadap bulan kunjungan pasien dalam setiap kluster. Ini membantu dalam mengidentifikasi pola kunjungan berdasarkan usia dan waktu.



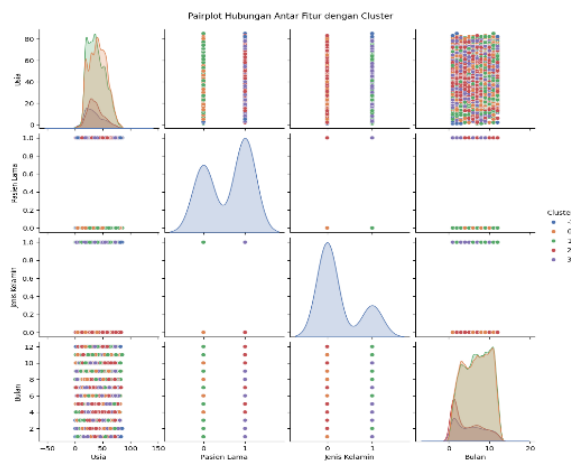
Gambar 3. Scatterplot Plot Usia vs Bulan dengan Kluster

- b. Scatter Plot Usia vs Jenis Kelamin dengan Kluster
- Menunjukkan distribusi usia berdasarkan jenis kelamin dalam setiap kluster, membantu dalam mengidentifikasi perbedaan distribusi gender dalam kluster.



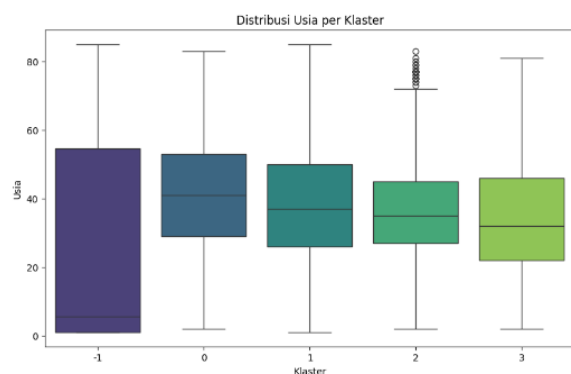
Gambar 4. Scatter Plot Usia vs Jenis Kelamin dengan Klaster

- c. Pair Plot untuk Melihat Hubungan Antar Fitur
Menunjukkan hubungan antara fitur-fitur seperti usia, status pasien (lama atau baru), jenis kelamin, dan bulan kunjungan dengan klaster yang terbentuk.



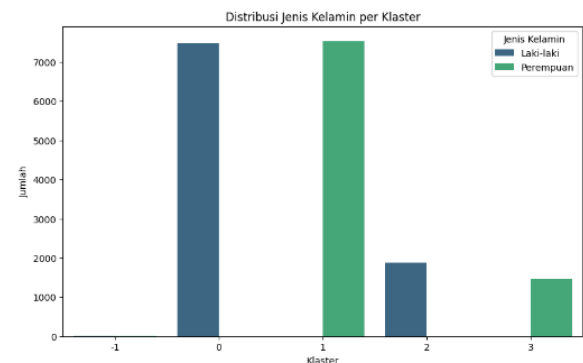
Gambar 5. Pair Plot untuk Melihat Hubungan Antar Fitur

- d. Distribusi Usia per Klaster
memberikan gambaran tentang median, kuartil, dan outlier dalam distribusi usia untuk setiap klaster. Ini membantu memahami apakah ada klaster tertentu yang mendominasi usia tertentu.



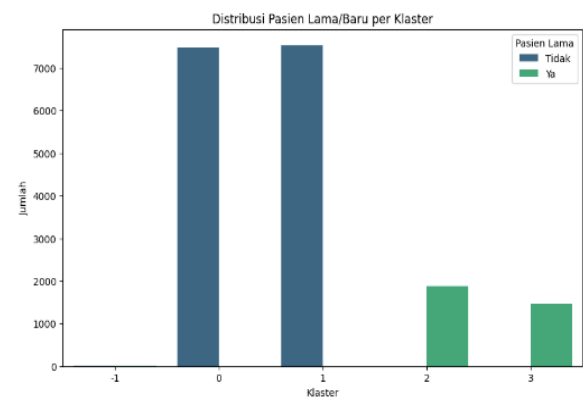
Gambar 6. Distribusi Usia per Klaster

- e. distribusi jenis kelamin per klaster
membantu untuk melihat distribusi jenis kelamin dalam setiap klaster dan apakah ada ketidakseimbangan jenis kelamin dalam klaster tertentu.



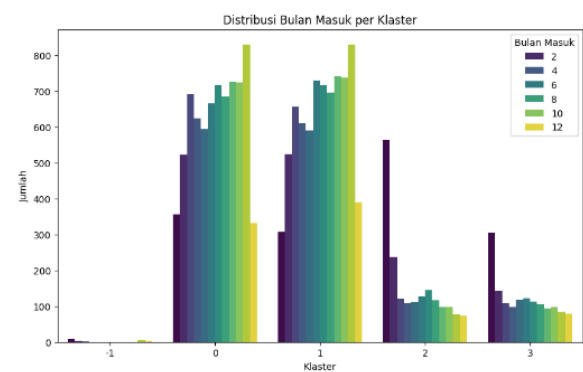
Gambar 7. distribusi jenis kelamin per klaster

- f. Distribusi pasien lama/baru per klaster
Membantu memahami bagaimana pasien lama dan baru terdistribusi dalam setiap klaster.



Gambar 8. Distribusi pasien lama/baru per klaster

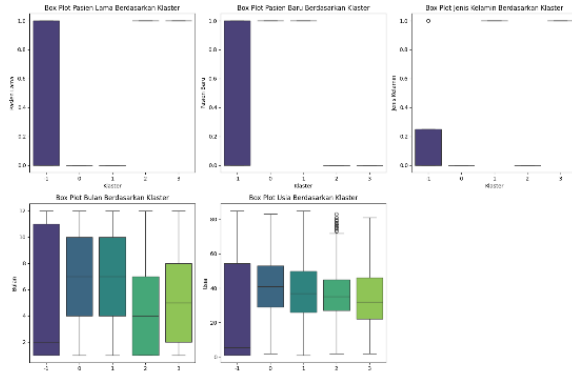
- g. Distribusi bulan masuk per klaster
memberikan wawasan tentang distribusi temporal pasien dalam setiap klaster dan apakah ada pola tertentu dalam waktu masuk pasien ke rumah sakit.



Gambar 9. Distribusi bulan masuk per klaster

- h. Boxplot untuk Setiap Fitur Berdasarkan Klaster
Boxplot ini menunjukkan distribusi dari masing-masing fitur (Pasien Lama, Pasien Baru, Jenis

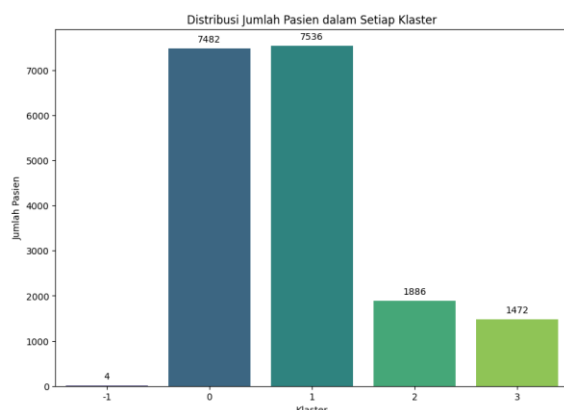
Kelamin, Bulan, Usia) untuk setiap kluster. Boxplot ini membantu untuk melihat distribusi dari masing-masing fitur dalam setiap kluster dan apakah ada fitur tertentu yang mendominasi kluster tertentu.



Gambar 10. Boxplot untuk Setiap Fitur Berdasarkan Kluster

4.3. Hasil Klasterisasi DBSCAN

Berdasarkan hasil penerapan algoritma DBSCAN pada data pasien rumah sakit jiwa kota kendari tahun 2022-2023, didapatkan enam kluster utama, termasuk satu kluster outlier. Berikut adalah rincian distribusi jumlah pasien dalam setiap kluster.



Gambar 11 Distribusi Jumlah Pasien dalam Setiap .Kluster

Hasil rincian distribusi jumlah pasien dalam setiap kluster, Cluster -1 (Outliers): Terdiri dari 4 pasien. Cluster 0: Kluster ini terdiri dari 7482 pasien. Cluster 1: Kluster ini merupakan yang terbesar dengan 7536 pasien. Cluster 2: Kluster ini terdiri dari 1886 pasien. Cluster 3: Terdiri dari 1472 pasien. Berikut adalah Analisis Karakteristik Tiap Kluster :

- Klaster -1 terdiri dari 4 pasien yang dianggap sebagai outliers. Dari jumlah ini, 58.33% adalah pasien lama dan 41.67% adalah pasien baru. Usia pasien dalam kluster ini sangat bervariasi dengan rata-rata usia 28 tahun dan rentang usia dari 1 hingga 85 tahun. Jenis kelamin pasien dalam kluster ini adalah 25% perempuan dan 75% laki-laki. Karakteristik utama dari kluster ini adalah

variasi usia yang sangat besar dengan mayoritas pasien laki-laki.

- Klaster 0 terdiri dari 7482 pasien yang semuanya adalah pasien baru. Tidak ada pasien lama dalam kluster ini. Rata-rata usia pasien dalam kluster ini adalah 41 tahun dengan rentang usia dari 2 hingga 83 tahun. Semua pasien dalam kluster ini adalah laki-laki. Karakteristik utama dari kluster ini adalah bahwa mereka semua merupakan pasien baru yang semuanya laki-laki dengan usia rata-rata yang lebih tua dibandingkan dengan kluster lainnya.
- Klaster 1 terdiri dari 7536 pasien yang semuanya adalah pasien baru, mirip dengan Klaster 0. Tidak ada pasien lama dalam kluster ini. Rata-rata usia pasien dalam kluster ini adalah 39 tahun dengan rentang usia dari 1 hingga 85 tahun. Semua pasien dalam kluster ini adalah perempuan. Karakteristik utama dari kluster ini adalah bahwa mereka semua merupakan pasien baru yang semuanya perempuan dengan usia rata-rata yang sedikit lebih muda dibandingkan dengan Klaster 0.
- Klaster 2 terdiri dari 1886 pasien yang semuanya adalah pasien lama. Tidak ada pasien baru dalam kluster ini. Rata-rata usia pasien dalam kluster ini adalah 37 tahun dengan rentang usia dari 2 hingga 83 tahun. Semua pasien dalam kluster ini adalah laki-laki. Karakteristik utama dari kluster ini adalah bahwa mereka semua merupakan pasien lama yang semuanya laki-laki dengan usia menengah.
- Klaster 3 terdiri dari 1472 pasien yang semuanya adalah pasien lama, mirip dengan Klaster 2. Tidak ada pasien baru dalam kluster ini. Rata-rata usia pasien dalam kluster ini adalah 35 tahun dengan rentang usia dari 2 hingga 81 tahun. Semua pasien dalam kluster ini adalah perempuan. Karakteristik utama dari kluster ini adalah bahwa mereka semua merupakan pasien lama yang semuanya perempuan dengan usia menengah..

4.4. Evaluasi Klasterisasi dengan Silhouette Score

Silhouette Score menghitung lebar berdasarkan jarak pada setiap titik atau data yang menjadi anggota suatu cluster, kemudian lebar diperoleh dengan menghitung rata-rata jarak total. Silhouette Score digunakan untuk mengukur seberapa baik kluster terbentuk. Nilai Silhouette Score berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan bahwa kluster lebih baik terbentuk..

Hasil evaluasi menunjukkan bahwa rata-rata Silhouette Score yang diperoleh adalah 0.365. Nilai ini mengindikasikan bahwa kualitas klasterisasi berada dalam kategori sedang, di mana data tidak terlalu terpisah dengan jelas tetapi masih memiliki struktur kluster yang cukup baik.

Plot Silhouette yang dihasilkan menunjukkan distribusi koefisien Silhouette untuk setiap kluster. Plot ini membantu memahami kualitas pengelompokan data, termasuk bagaimana data dalam

setiap kluster dikelompokkan dan apakah ada kluster yang memiliki nilai negatif (indikasi bahwa beberapa data lebih dekat ke kluster lain daripada ke klasternya sendiri). Meskipun nilai Silhouette Score belum mencapai hasil optimal, analisis lebih lanjut terhadap parameter model atau pendekatan klasterisasi dapat dilakukan untuk meningkatkan kualitas klasterisasi.

5. KESIMPULAN DAN SARAN

Hasil penerapan algoritma DBSCAN pada data pasien Rumah Sakit Jiwa Kota Kendari tahun 2022-2023 mengidentifikasi enam kluster, termasuk satu kluster outlier. Kluster -1 (outliers) terdiri dari 4 pasien dengan variasi usia besar (1-85 tahun), didominasi pasien laki-laki (75%) dengan rata-rata usia 28 tahun. Kluster 0 mencakup 7482 pasien baru laki-laki dengan rata-rata usia 41 tahun, sedangkan kluster 1 terdiri dari 7536 pasien baru perempuan dengan rata-rata usia 39 tahun. Kluster 2 berisi 1886 pasien lama laki-laki dengan rata-rata usia 37 tahun, sementara kluster 3 mencakup 1472 pasien lama perempuan dengan rata-rata usia 35 tahun. Rata-rata Silhouette Score sebesar 0.365 menunjukkan kualitas klasterisasi sedang, dengan struktur kluster yang cukup baik meskipun data tidak sepenuhnya terpisah dengan jelas. Penyesuaian parameter model atau metode alternatif dapat meningkatkan kualitas klasterisasi. Hasil ini menyarankan perlunya perhatian khusus kepada pasien outliers karena karakteristiknya yang sangat bervariasi. Program layanan terpisah sebaiknya dikembangkan untuk pasien baru dan pasien lama, dengan menyesuaikan perawatan berdasarkan usia dan gender.

DAFTAR PUSTAKA

- [1] Y. Diana *et al.*, “Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang,” *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 7, no. 1, pp. 97–103, 2023.
- [2] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, “Pengelompokkan Data Runtun Waktu menggunakan Analisis Cluster (Studi Kasus: Nilai Ekspor Komoditi Migas dan Nonmigas Provinsi Kalimantan Timur Periode Januari 2000-Desember 2016),” *J. EKSPONENSIAL*, vol. 11, no. 1, pp. 29–38, 2020.
- [3] W. Rohalidyawati, R. Rahmawati, and M. Mustafid, “Segmentasi Pelanggan E-Money Dengan Menggunakan Algoritma Dbscan (Density Based Spatial Clustering Applications With Noise) Di Provinsi Dki Jakarta,” *J. Gaussian*, vol. 9, no. 2, pp. 162–169, 2020, doi: 10.14710/j.gauss.v9i2.27818.
- [4] E. Haerani, E. Budianita, A. Nazir, and W. Mahesa, “Penerapan K-Means Clustering Pada Data Obat/Alkes di Apotik RSUD Selasih,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, vol. 1, no. 1, pp. 220–229, 2023.
- [5] Y. Pratiwi and M. Mulyawan, “Implementasi Algoritma K-Means untuk Menentukan Angka Harapan Hidup berdasarkan Tingkat Provinsi,” *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 284–294, 2023, doi: 10.56211/blendsains.v1i4.233.
- [6] E. A. Fadlilah, “Identifikasi Anomali Data Akademik Menggunakan Dbscan Outlier Detection,” *Pros. Sains Nas. dan Teknol.*, vol. 12, no. 1, p. 336, 2022, doi: 10.36499/psnst.v12i1.7012.
- [7] A. Kristianto, “Implementasi DBSCAN dalam Clustering Data Minat Mahasiswa Setelah Pandemi Covid19,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 426–431, 2022, doi: 10.24002/konstelasi.v2i2.5638.
- [8] I. A. Nur Afifah and H. Nurdyanto, “Data Mining Clustering Dalam Pengelompokan Buku Perpustakaan Menggunakan Algoritma K-Means,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 8, no. 3, pp. 802–814, 2023, doi: 10.29100/jupi.v8i3.3891.
- [9] D. Ariyanti and K. Iswardani, “Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes,” *J. IKRA-ITH Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- [10] S. H. Hastuti, A. Septiani, H. Hendrayani, and W. P. Nurmayanti, “Penerapan Metode OPTICS dan ST-DBSCAN untuk Klasterisasi Data Kesehatan,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 252–261, 2024, doi: 10.29408/edumatic.v8i1.25765.