

## Klasterisasi Data Rekam Medis Pasien Menggunakan Metode *K-Means Clustering* Di Rumah Sakit Widodo Ngawi

Harfina Dilawati<sup>1</sup>, Heru Widiyanto<sup>2\*</sup>, Agustinus Kuswiadji<sup>3</sup>

<sup>1</sup>D3 Perekam dan Informasi Kesehatan, STIKES Bhakti Husada Mulia, harfinadilawati@gmail.com

<sup>2\*</sup>D3 Perekam dan Informasi Kesehatan, STIKES Bhakti Husada Mulia, widiyanto.masher@gmail.com

<sup>3</sup>D3 Perekam dan Informasi Kesehatan, STIKES Bhakti Husada Mulia, agustinuskuswiadji@gmail.com

### Keywords:

Medical Records,  
Clustering,  
Data Mining,  
K-Means Clustering,  
RStudio,

### ABSTRACT

Medical record management is the most vital aspect of hospital operations. Medical records serve as a crucial source of information in patient healthcare services. Through data mining, knowledge extraction and data analysis can be carried out to find relationships, data structures, patterns, and regularities between data. The purpose of this study is to cluster using the RStudio application with the K-means clustering technique and identify disease prevalence patterns in Ngawi Regency. The methodology used is quantitative descriptive using secondary data, samples taken from hospitalization data from October to December 2023 totaling 3171 and divided into six variables, namely gender, age, sub-district, diagnosis, length of treatment and payment method. There are four clustering results, Cluster 1 amounted to 524 patients (19%), Cluster 2 amounted to 831 patients (30%), Cluster 3 amounted to 940 patients (33%) and Cluster 4 amounted to 512 patients (18%). Based on visits from each sub-district in Ngawi Regency, the total number of inpatient disease cases is fever (74), gastrointestinal diseases (64), stroke (43), and degenerative diseases such as diabetes (64), heart disease (7), neoplasms (6), asthma (4), systemic diseases (3), and external diseases (10). The study's findings can serve as a foundation for statistical management to inform decisions aimed at enhancing services and inpatient facilities.

### Kata Kunci

Rekam Medis,  
Klasterisasi,  
Data Mining,  
K-Means Clustering,  
RStudio,

### ABSTRAK

Manajemen data jejak medis merupakan aspek paling vital dalam operasional rumah sakit. Data rekam medis berfungsi sebagai sumber informasi krusial dalam pelayanan kesehatan bagi pasien. Dengan menggunakan data mining, peneliti dapat mengekstrak pengetahuan dan menganalisis data untuk mengidentifikasi korelasi, struktur, pola, dan keteraturan di antara data tersebut. Tujuan penelitian ini melakukan clusterisasi menggunakan aplikasi RSudio dengan teknik K-means clustering serta mengidentifikasi pola prevalensi penyakit di Kabupaten Ngawi. Metodologi studi ini ialah deskriptif kuantitatif melalui data sekunder, sampel diambil dari data rawat inap bulan Oktober hingga Desember 2023 berjumlah 3171 dan dibagi menjadi enam variabel yaitu jenis kelamin, usia, kecamatan, diagnosa, lama dirawat dan cara bayar. Terdapat empat hasil clusterisasi, Cluster 1 berjumlah 524 pasien (19%), cluster 2 berjumlah 831 pasien (30%), Cluster 3 berjumlah 940 pasien (33%) dan cluster 4 berjumlah 512 pasien (18%). Berdasarkan kunjungan dari tiap kecamatan di Kabupaten Ngawi, total kasus penyakit terbanyak pasien rawat inap adalah demam (74), penyakit saluran cerna (64), stroke (43), dan penyakit degeneratif seperti diabetes (64) penyakit jantung (7), neoplasma (6), asma (4), penyakit sistemik (3), dan penyakit akibat luar (10). Hasil penelitian dapat digunakan sebagai dasar pengolahan statistik untuk mengambil keputusan dalam peningkatan pelayanan dan fasilitas rawat inap.

### Korespondensi Penulis:

Heru Widiyanto,  
STIKES Bhakti Husada Mulia,  
Jl. Taman Praja No.25, Mojorejo, Kec. Taman, Kota Madiun  
Telepon: +6285231226357  
Email: widiyanto.masher@gmail.com

Submitted: 05-08-2024; Accepted: 23-09-2024;  
Published: 29-09-2024

*Copyright* (c) 2024 The Author (s) This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

## 1. PENDAHULUAN

Dokumen rekam medis mencakup dokumen terkait tindakan medis, pengobatan, pemeriksaan, identitas pasien, dan layanan lainnya bagi pasien [1]. Pengelolaan data tersebut menjadi bagian terpenting dalam operasional rumah sakit. Data rekam medis sebagai informasi mempunyai peran krusial dalam melayani pasien terkait hal-hal medis. Administrasi data rekam medis yang efektif sangat bermanfaat dalam memfasilitasi pengambilan keputusan. Namun demikian, rumah sakit masih mengalami kesulitan dalam mengelola data rekam medis secara efektif [2]. Rekam medis yang terakumulasi menjadi tantangan serius karena pengelolaan yang kurang efisien menyebabkan kurangnya pemanfaatan data riwayat pasien pada kegiatan pelaporan, yang pada gilirannya mempengaruhi kualitas pelayanan. Dengan tumpukan arsip yang sulit dikelola, rumah sakit sering kali hanya fokus pada kondisi pasien saat ini dan mengabaikan informasi penting dari riwayat rekam medis.

Pertumbuhan teknologi informasi telah memfasilitasi penyimpanan data dalam jaringan komputer, sehingga menghasilkan pembentukan sistem basis data yang luas [3]. Data mining menjelaskan proses melalui teknologi AI, matematika, dan statistik. Dengan menggunakan data mining, peneliti dapat mengekstrak pengetahuan dan menganalisis data untuk mengidentifikasi hubungan, struktur, pola, dan keteraturan di antara data [4]. Data mining merupakan sebuah proses yang menggunakan teknik kecerdasan buatan, dan *machine learning* yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang berkaitan dengan dengan berbagai *database* yang besar [5]. Salah satu fungsi data mining adalah clustering. Clustering merupakan tehnik pengelompokan record pada basis data berdasarkan kriteria tertentu. Hasil clustering menggambarkan tentang apa yang terjadi pada basis data [6]. Proses *data mining* merupakan proses yang terdiri dari pengumpulan data, ekstraksi data, analisis data, dan statistika. Proses ini juga umum dikenal dengan *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lainnya [7].

Satu dari sekian banyak pendekatan dalam data mining yang berfungsi guna melakukan analisis data dan memperoleh informasi yang relevan disebut *clustering*. Pendekatan ini berfokus pada pengelompokan data atau objek ke dalam beberapa kategori [8]. Clustering adalah teknik dalam analisis data guna melakukan pengelompokan data ke dalam grup-grup yang memiliki kemiripan tertentu. Salah satu metode populer dalam clustering adalah K-Means. Metode ini termasuk dalam kategori data clustering non-hierarki, yang berarti ia tidak membangun hierarki atau struktur bertingkat dalam proses pengelompokan. Sebaliknya, K-Means bekerja dengan mempartisi dataset ke dalam sejumlah kelompok atau cluster berdasarkan kesamaan fitur [9].

Penelitian terdahulu terkait dengan topik penelitian ini yaitu sebagai berikut : penggunaan *data mining* dengan metode *k-means clustering*, dapat mengidentifikasi informasi rekam medis bersumber pada wilayah, jenis penyakit dan umur di Poli klinik PT.Inecda [1]. Dengan Metode klasterisasi *data mining* dapat menjadi metode yang diharapkan dapat mempermudah dalam menghitung data penyakit pasien pada Puskesmas Lubuk Alung [10]. penggunaan *data mining* dengan metode *k-means clustering*, dapat menghasilkan informasi baru yaitu klasterisasi data rekam medis pada kecamatan, diagnosa penyakit, usia dan jenis kelamin berdasarkan rekam medis di rumah Sakit Anwar Medika [11].

Berdasarkan penjelasan di atas, peneliti hendak melakukan studi dengan melakukan pengelompokan data rekam medis pasien berdasarkan data variabel diagnosa penyakit, kecamatan, usia, jenis kelamin, lama dirawat dan cara bayar menggunakan aplikasi RStudio melalui *K-Means Clustering*. Diharapkan dengan menerapkan data mining Algoritma *K-Means Clustering* dapat memberikan informasi baru dari hasil klasterisasi jejak medis pasien sehingga dapat meningkatkan kinerja pada rumah sakit. Manfaat lainnya dari penelitian ini dapat membaca pola penyebaran penyakit, agar rumah sakit mempunyai gambaran pola penyebaran penyakit berdasarkan atribut yang ada pada rekam medis pasien.

## 2. METODE PENELITIAN

### 2.1 Metodologi Penelitian

Studi ini menerapkan metode kuantitatif deskriptif, data dikumpulkan dan dianalisis secara numerik untuk memberikan gambaran yang akurat tentang objek penelitian. Pendekatan cross sectional memungkinkan pengumpulan data pada satu titik waktu tertentu, sehingga memberikan gambaran sekilas dari situasi yang sedang diteliti [12]. Proses penelitian dimulai dengan mengumpulkan data rekam medis. Setelah data terkumpul, tahap berikutnya adalah pembersihan data, di mana data yang tidak lengkap akan disingkirkan dari analisis. Tahap berikutnya adalah preprosesing data, yaitu memasukkan data tersebut ke dalam spreadsheet Excel. Langkah ini bertujuan untuk memudahkan perangkat lunak dalam melakukan perhitungan dan analisis data. Setelah data diproses dan dimasukkan ke dalam Excel, langkah berikutnya adalah clustering atau pengelompokan data. Terakhir, data dianalisis untuk menemukan pola, tren, atau informasi yang berguna bagi penelitian [10].

### 2.2 Data Mining

Cabang ilmu yang berfokus pada teknik-teknik untuk menggali pengetahuan atau mengidentifikasi pola dari kumpulan data. Secara umum, disiplin ilmu ini melibatkan metode-metode pengolahan data yang dirancang untuk mengungkap hubungan yang tidak tampak dalam suatu data [13].

### 2.3 Metode K-Means Clustering

Metode yang dapat dimanfaatkan untuk mengelompokkan atau cluster data adalah K Means. Dalam metode ini, data yang memiliki kesamaan atau kemiripan dikelompokkan dalam satu kelompok yang sama, sementara data yang berbeda atau memiliki kriteria yang variatif akan dipisahkan ke dalam beberapa cluster yang terpisah [14].

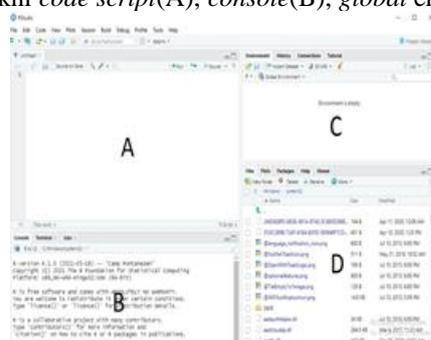


Gambar 1. Tahapan Algoritma K- Means

Dari Gambar 1 dapat dilihat diagram alir K-Means dimulai dengan menentukan banyaknya cluster. Setelah nilai ditentukan, langkah selanjutnya adalah menentukan pusat cluster, dan dilanjutkan dengan menghitung jarak setiap objek setiap pusat cluster. Jarak antara tiap titik data yang ada dan tiap centroid akan dihitung menggunakan analisis Euclidean hingga ditentukan jarak terpendek antara setiap titik data dan centroid. Selanjutnya dilakukan pengelompokan objek berdasarkan jarak minimum terhadap pusat cluster. Pusat cluster tersebut kemudian untuk sementara dijadikan pusat cluster, atau centroid, mean. Jika masih terdapat objek yang harus dipindah ke cluster yang lain, maka proses diulang kembali, tetapi jika tidak, maka proses selesai [15].

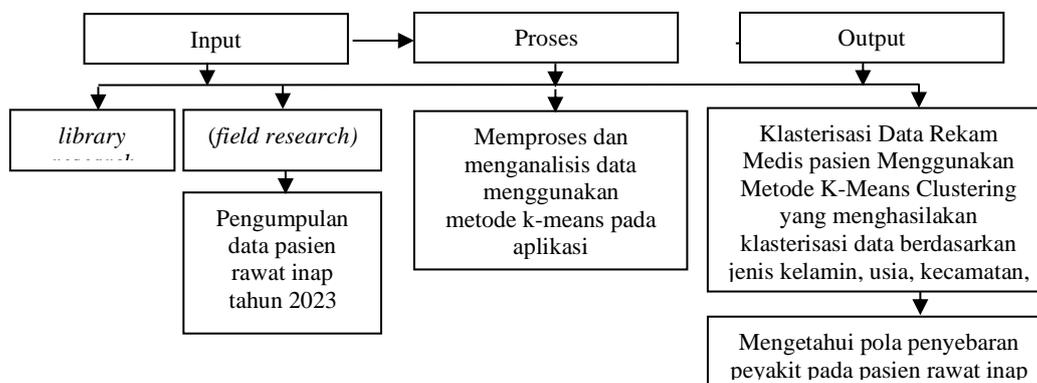
**2.4 Aplikasi RStudio**

RStudio adalah bagian dari IDE (*Integrated Development Environment*), Pada tahun 2013, RStudio meraih posisi teratas dalam pemilihan KD Nuggets 2013, diakui sebagai salah satu “Top Language for Analytics”, untuk data mining dan penelitian ilmiah. Dapat dilihat pada Gambar 2, antarmuka RStudio dibagi menjadi empat bagian utama, masing-masing dengan fungsi yang spesifik, yakni *code script*(A), *console*(B), *global environment*(C), dan *menu file*(D).



Gambar 2. Tampilan Awal Jendela Rstudio

**2.5 Diagram Aktivitas Kerja Penelitian**



Gambar 3. Aktivitas Kerja Penelitian

Gambar 3 di atas menunjukkan input pada penelitian ini adalah pencarian informasi dari penelitian terdahulu serta studi pendahuluan dengan mengidentifikasi kunjungan pasien rawat Inap di Rumah Sakit Widodo Ngawi. Disini peneliti juga melakukan tahapan pengumpulan data rekam medis yang nantinya data tersebut akan diolah dalam tahapan proses. Dalam tahapan proses menerapkan teknik data mining melalui K-Means Clustering, serta peneliti juga melaksanakan proses analisis data hasil pengklasteran. Output dari penelitian ini adalah mengidentifikasi karakteristik data jejak medis pasien dalam database. Tujuannya adalah untuk menganalisis dan memproses data tersebut agar dapat menemukan informasi baru yang relevan.

### 3. HASIL DAN ANALISIS

#### 3.1 Alur Proses Pengelolaan Data Rekam Medis Pasien Rawat Inap di Rumah Sakit Widodo Ngawi

Berdasarkan hasil wawancara antara peneliti dengan responden yaitu Kepala Unit Rekam medis sebagai informan 1 dan petugas pelaporan sebagai informan 2 di Rumah Sakit Widodo Ngawi Rekam Medis sudah dilakukan pengelompokan berdasarkan kategori tertentu. Untuk proses pengelolaan data Rekam Medis menggunakan SIMRS dan juga secara manual menggunakan Microsoft Excel yang di lakukan oleh petugas pelaporan pada unit rekam medis. Data yang sudah di kategorikan dan diolah dimanfaatkan rumah sakit untuk proses pelaporan internal dan juga pelaporan eksternal rumah sakit. Studi ini menerapkan data sekunder dari database rekam medis yang di ambil berjumlah 3171 data. Data yang di peroleh dari triwulan ke empat bulan Oktober sampai Desember pasien rawat inap di Rumah Sakit Widodo Ngawi tahun 2023. Data yang diterima dalam bentuk tabel pada Microsoft Excel.

#### 3.2 Proses Pengelolaan Data Rekam Medis

##### 3.2.1. Identifikasi Data

Data mempunyai jumlah baris sebanyak 3171 dengan 6 variabel yaitu diagnosa penyakit, kecamatan, usia, jenis kelamin, lama dirawat dan cara bayar. Identifikasi data tersebut dapat dilihat pada Gambar 4 di bawah ini

	JENIS KELAMIN	UMUR	ALAMAT	DIAGNOSA	LD	CARA BAYAR
1	PEREMPUAN	BALITA	BOJONEGORO	GEA	5	UMUM
2	LAKI-LAKI	BALITA	PARON	COR	3	UMUM
3	PEREMPUAN	MANULA	PARON	Stroke	7	UMUM
4	PEREMPUAN	BALITA	BRINGIN	GEA	3	UMUM
5	PEREMPUAN	BALITA	KARANGANYAR	COR	4	UMUM
6	PEREMPUAN	BALITA	BRINGIN	Stroke Infark	7	UMUM
7	LAKI-LAKI	BALITA	SRAGEN	CF Manus Dx	4	UMUM
8	LAKI-LAKI	BALITA	PARON	Of Manus Dx	4	UMUM

Gambar 4. Proses Identifikasi Data Awal

##### 3.2.2. Library

Untuk mengolah data mining dan melakukan pemodelan K-Prototypes Clustering, beberapa pustaka (libraries) sangat diperlukan seperti penggunaan *library readxl* untuk membaca Microsoft Excel xlsx dari dalam R, *library ggplot2* untuk visualisasi data, *library dplyr* untuk mengolah data dalam bahasa intuitif serta manipulasi data, dan *library clusteMixType* untuk proses klasterisasi.

#### 3.2.3. Tahapan Pada Data Mining

##### a. Cek Missing values

Pengecekan *missing values* atau *data selection* merupakan tahapan untuk menghapus *noise* data yang bisa mengganggu saat penggalian data dalam *Knowledge Discovery In Database (KDD)*. Diawali dengan data hasil pilih yang hendak digunakan dalam proses data mining disimpan dalam suatu berkas.

##### b. Melakukan Data Normalization

Normasilsasi data digunakan dengan melakukan scaling pada data jenis data *numeric*. Pada tahap normalisasi ini terdapat penskalaan data pada *dataframe* untuk melihat nilai-nilai menggunakan Min-Max Scaler dengan rentang [0, 1].

##### c. Data Transformation

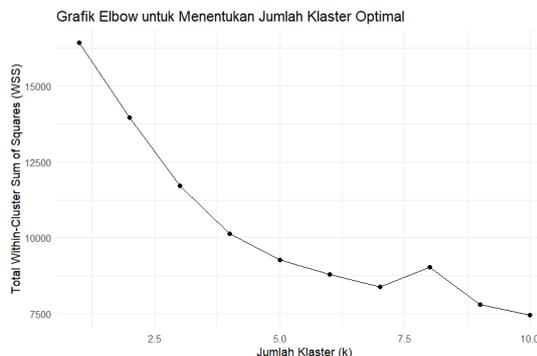
Data transformation adalah proses mengubah atau mengkonveksi data dari satu bentuk atau format ke bentuk yang lain untuk mudah analisis atau pemrosesan, umumnya dilakukan dengan mengubah variabel kategori menjadi faktor.

##### d. Menentukan Jumlah Cluster K-Prototype Clustering

Jumlah cluster ditentukan menggunakan *elbow method*, yaitu jumlah perbandingan antara WWS (*Total within cluster sum of square*) dengan jumlah klaster. Dimana grafik yang dihasilkan akan membantu menentukan jumlah *cluster* yang tepat untuk *dataset*.

```
>
> # Buat data frame untuk plot
> wss_data <- data.frame(k = k.values, wss = wss_values)
>
> # Plot grafik elbow
> ggplot(wss_data, aes(x = k, y = wss)) +
+   geom_line() +
+   geom_point() +
+   labs(title = "Grafik Elbow untuk Menentukan Jumlah Kluster Optimal",
+         x = "Jumlah Kluster (k)",
+         y = "Total Within-Cluster Sum of Squares (WSS)") +
+   theme_minimal()
> |
```

Gambar 5. Tahap Klasterisasi



Gambar 6. Grafik Elbow

Dari Gambar 6 di atas diketahui grafik elbow terlihat bahwa jumlah kluster paling optimal adalah k=4. Yaitu dengan menginisialisasi awal nilai k kemudian nilai k yang mengalami penurunan secara drastic, yang mana penurunan inersia mulai terjadi perlambatan, melakukan pembentukan grafik menyerupai bentuk “siku” sehingga dapat ditetapkan sebagai k optimal.

**e. Membuat Centroid**

Pada tahap ini *centroid* diperlukan guna menentukan posisi pusat dari setiap cluster yang telah terbentuk selama proses pembuatan model K-Prototypes. Untuk menentukan hasil clustering maka diperlukan fungsi `enumerate()`. Untuk dapat mengakses indeks dan koordinat pusat dari *cluster\_centers*. Dari kode tersebut didapatkan hasil centroid untuk masing-masing kluster. Adapun hasil centroid kluster dapat dilihat pada Gambar 7 di bawah ini.

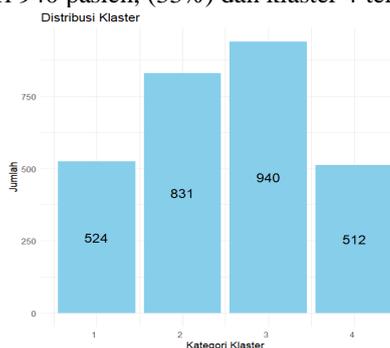
	JENIS KELAMIN	UMUR[, 1]	ALAMAT[, 1]	DIAGNOSA	LAMA DIRAWAT[, 1]	CARA BAYAR[, 1]	cluster
1	2	0.4954321	0.1912519	I63.9	-0.24677858	-0.4584346	1.365338
2	1	0.6104233	0.3136706	I63.9	1.47814395	0.3779785	2.960630
3	1	-0.7665186	0.4374796	O33.9	-0.30324904	0.4789998	4.000000
4	1	-0.5809931	-1.7045208	A97.0	-0.05176824	0.1034964	3.614634

Gambar 7. Hasil Klasterisasi

Dari pusat centroid ini, peneliti dapat menganalisis, setiap centroid menjadi perwakilan dari sebuah kelompok dalam hasil pengelompokan. Koordinat centroid memberikan gambaran tentang pusat atau representasi mean dari semua sampel dalam kelompok itu.

**f. Memvisualisasi Hasil Analisis dan Pembagian Data Cluster**

Visualisasi hasil kluster dilakukan menggunakan bar chart agar terlihat jelas perbedaan jumlah anggota dalam masing-masing kluster. Dari gambar dapat dilihat bahwa kluster 1 terdiri dari 524 pasien (19%), kluster 2 terdiri dari 831 pasien (30%), kluster 3 terdiri dari 940 pasien (33%) dan kluster 4 terdiri dari 512 pasien (18%).



Gambar 8. Hasil Visualisasi Klasterisasi

**g. Hasil Analisa Data Setiap Cluster**

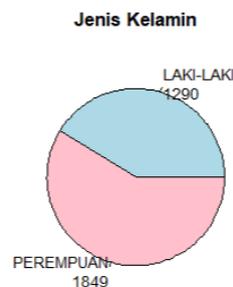
Tabel 1. Hasil Klasterisasi

Keterangan	Klaster 1	Klster 2	Klaster 3	Klaster 4
Jumlah data	524	831	940	512
Jenis kelamin	Perempuan = 280 Laki-laki = 244	Perempuan = 659 Laki-laki = 172	Perempuan = 421 Laki-laki = 519	Perempuan = 321 Laki-laki = 191
Alamat	Geneng, Pitu, Kasreman, Sine, Ngrambe, Jogorogo, Karangjati, Paron, Padas, Ngawi, Bringin, Pangkur, Kendal, Mantingan, Kedunggalar, Gerih	Kwadungan, Kasreman, Sine, Geneng, Mantingan, Padas, Jogorogo, Pitu, Karangjati, Pangkur, Kendal, Gerih, Ngrambe, Bringin, Widodaren	Karanganyar, Kwadungan, Sine, Pitu, Geneng, Gerih, Ngrambe, Widodaren, Kasreman, Bringin, Jogorogo, Kendal, Padas, Pangkur, Mantingan, Kedunggalar, Ngawi, Paron	Ngawi
Lama rawat	1-11 hari	1-15 hari	1-53 hari	1-11 hari
Cara bayar	Umum= 524 BPJS = 0	Umum= 0 BPJS = 831	Umum= 0 BPJS = 940	Umum= 0 BPJS = 512

**3.2.4. Proses Pengelolahan Kontingensi antar Variable**

**a. Pengelompokan Variable Jenis Kelamin**

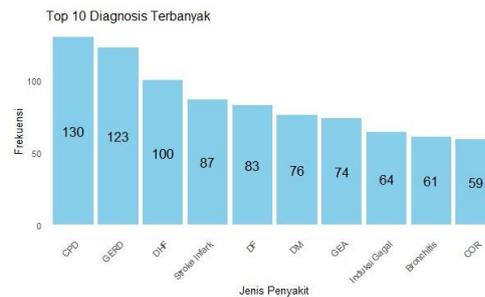
Faktor jenis kelamin dan gender merupakan salah satu faktor yang berpengaruh terhadap pemanfaatan fasilitas pelayanan kesehatan. Kecenderungan pemilihan oleh pasien laki-laki dan perempuan dapat dipengaruhi oleh kondisi ekonomi, fasilitas kesehatan, pendidikan, kenyamanan, dan lain-lain. Selain itu data pelaporan jenis kelamin di rumah sakit dapat digunakan untuk tabulasi jumlah kunjungan pasien pada bangsal berdasarkan jenis kelaminnya. Dari hasil pengelompokan jumlah kunjungan pasien rawat inap triwulan ke 4 tahun 2023 rasio jumlah kelamin perempuan lebih banyak daripada laki-laki dengan jumlah 1849 dan 1290 dari total kunjungan 3 139 pasien. Hasil Pengelompokan variabel jenis kelamin dapat dilihat pada Gambar 9 di bawah ini.



Gambar 9. Hasil Pengelompokan Variabel Jenis Kelamin

**b. Pengelompokan 10 Besar Diagnosa Penyakit**

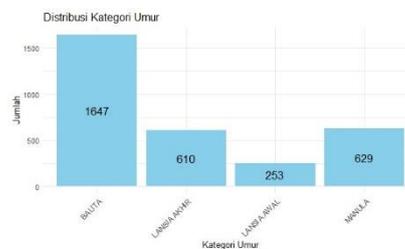
Pembangunan kesehatan bertujuan untuk meningkatkan kesadaran, kemauan, kemampuan hidup sehat bagi setiap orang agar terwujud derajat kesehatan yang optimal. Kondisi kesehatan masyarakat di suatu wilayah dapat dilihat dari jumlah kasus penyakit yang berkunjung ke Rumah Sakit. Dari data hasil kunjungan pasien rawat inap di Rumah Sakit Widodo Ngawi pada triwulan ke 4 tahun 2023 dapat dilihat pada gambar di bawah ini. Pada umumnya pelaporan 10 besar penyakit digunakan untuk pelaporan RL5 Data Bulanan yang terkhusus pada RL5.3. Data 10 Besar Penyakit Rawat Inap. Grafik pengelompokan 10 besar penyakit dapat dilihat pada Gambar 10 di bawah ini.



Gambar 10. Hasil Pengelompokan Diagnosa Penyakit

**c. Pengelompokan Usia Berdasarkan Kategori**

Dalam konsep dasar epidemiologi yang memberikan gambaran tentang hubungan antara tiga faktor utama yang berperan dalam terjadinya suatu penyakit dan masalah penyakit lainnya salah satunya adalah *host*. Faktor *host* adalah semua faktor yang terdapat dalam diri manusia yang dapat mempengaruhi timbulnya penyakit atau mempengaruhi perjalanan suatu penyakit, salah satu faktor *host* tersebut adalah usia. Seperti terlihat pada Gambar 11, Usia dapat dikategorikan dalam beberapa rentang yaitu balita (0-5), anak-anak (5-11), remaja awal (12-16), remaja akhir(17-25), dewasa awal(26-35), dewasa akhir(36-45), lansia awal (45-55) lasia akhir ( 56-65) dan manula (>65). Hasil pengelompokan usia berdasarkan Kategori dapat dilihat pada Gambar 11 di bawah ini.



Gambar 11. Hasil Pengelompokan Kategori Usia

**d. Pengelompokan Kontigensi Antara Usia dengan Jenis Kelamin, Diagnose Penyakit, Lama Dirawat dan Jenis Pembiayaannya**

...1 (double)	Var1 (character)	Var2 (character)	Var3 (character)	Var4 (character)	Var5 (character)
1	LAKI-LAKI	BALITA	BALI	Abdominal Pain	BPJS
2	PEREMPUAN	BALITA	BALI	Abdominal Pain	BPJS
3	LAKI-LAKI	LANSIA AKHIR	BALI	Abdominal Pain	BPJS
4	PEREMPUAN	LANSIA AKHIR	BALI	Abdominal Pain	BPJS
5	LAKI-LAKI	LANSIA AWAL	BALI	Abdominal Pain	BPJS
6	PEREMPUAN	LANSIA AWAL	BALI	Abdominal Pain	BPJS
7	LAKI-LAKI	MANULA	BALI	Abdominal Pain	BPJS
8	PEREMPUAN	MANULA	BALI	Abdominal Pain	BPJS
9	LAKI-LAKI	BALITA	BALIKAPAPAN	Abdominal Pain	BPJS
10	PEREMPUAN	BALITA	BALIKAPAPAN	Abdominal Pain	BPJS
11	LAKI-LAKI	LANSIA AKHIR	BALIKAPAPAN	Abdominal Pain	BPJS

Gambar 12. Tahap Pengelompokan Kontigensi Antar Variabel

Gambar 12 diatas menjelaskan tentang pengelompokan usia 0-5 tahun termasuk usia balita, dengan jumlah 299, jenis kelamin laki laki 156 dan perempuan 143, diagnosa penyakit terbanyak yaitu terkait sistem pencernaan 91 kasus, bronchitis 48 kasus dan demam 24 kasus dengan rata rata lama dirawat 1-26 hari dengan kategori pasien umum 102 dan pasien BPJS 197. Pada pengelompokan usia balita asal kecamatan terbanyak yaitu 68 dari Ngawi, 32 dari Paron, dan 20 dari Pitu.

Pengelompokan usia 5-11 tahun termasuk usia anak-anak, dengan jumlah 115, jenis kelamin laki laki 60 dan perempuan 55, diagnosa penyakit terbanyak yaitu terkait sistem pencernaan 21 kasus, bronchitis 12 kasus dan demam 11 kasus dengan rata rata lama dirawat 1-9 hari dengan kategori pasien umum 31 dan pasien BPJS 84. Pada pengelompokan usia anak-anak asal kecamatan terbanyak yaitu 36 dari Ngawi, 13 dari Paron, dan 9 dari Kedungalar.

Pengelompokan usia 12-16 tahun termasuk usia remaja awal, dengan jumlah 299, jenis kelamin laki laki 156 dan perempuan 143, diagnosa penyakit terbanyak yaitu patah tulang dan cedera 45 kasus, demam 29 kasus dan sistem pencernaan 14 kasus dengan rata rata lama dirawat 1-10 hari dengan kategori pasien umum 102 dan pasien BPJS 197. Pada pengelompokan usia remaja awal asal kecamatan terbanyak yaitu 26 dari Ngawi, 18 dari Paron, dan 11 dari Widodaren.

Pengelompokan usia 17-25 tahun termasuk usia remaja akhir, dengan jumlah 305, jenis kelamin laki laki 72 dan perempuan 233, diagnosa penyakit terbanyak yaitu persalinan 95 kasus, patah tulang dan cedera 47 kasus dan demam 30 kasus dengan rata rata lama dirawat 1-11 hari dengan kategori pasien umum 64 dan pasien BPJS 241. Pada pengelompokan usia remaja akhir asal kecamatan terbanyak yaitu 44 dari Ngawi, 32 dari Geneng, dan 27 dari Kedunggalar.

Pengelompokan usia 26-35 tahun termasuk usia dewasa awal, dengan jumlah 459, jenis kelamin laki laki 90 dan perempuan 369, diagnosa penyakit terbanyak yaitu terkait persalinan 222 kasus, sistem pencernaan 22 kasus dan demam 13 kasus dengan rata rata lama dirawat 1-12 hari dengan kategori pasien umum 87 dan pasien BPJS 372. Pada pengelompokan usia dewasa awal asal kecamatan terbanyak yaitu 102 dari Ngawi, 55 dari Paron, dan 45 dari Geneng

Pengelompokan usia 36-45 tahun termasuk usia dewasa akhir, dengan jumlah 325, jenis kelamin laki laki 103 dan perempuan 222, diagnosa penyakit terbanyak yaitu terkait persalinan 52 kasus, demam 35 kasus dan sistem pencernaan 28 kasus dengan rata rata lama dirawat 1-13 hari dengan kategori pasien umum 60 dan pasien BPJS 265. Pada pengelompokan usia dewasa akhir asal kecamatan terbanyak yaitu 61 dari Ngawi, 37 dari Paron, dan 32 dari Padas.

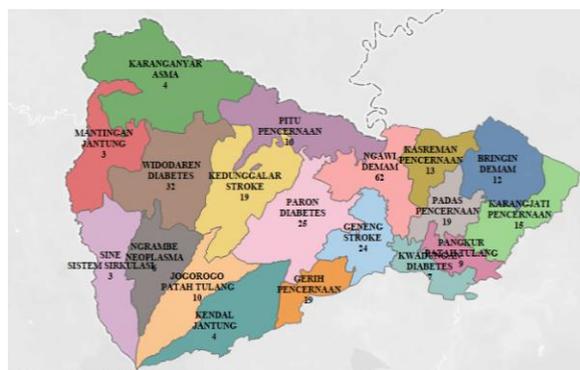
Pengelompokan usia 45-55 tahun termasuk usia lansia awal, dengan jumlah 317, jenis kelamin laki laki 172 dan perempuan 145, diagnosa penyakit terbanyak yaitu terkait diabetes 38 kasus, sistem pencernaan 28 kasus dan stroke 24 kasus dengan rata rata lama dirawat 1-8 hari dengan kategori pasien umum 56 dan pasien BPJS 261. Pada pengelompokan usia lansia awal asal kecamatan terbanyak yaitu 88 dari Ngawi, 43 dari Geneng, dan 40 dari Paron.

Pengelompokan usia 56-65 tahun termasuk usia lansia akhir, Dengan jumlah 577, jenis kelamin laki laki 284 dan perempuan 293, diagnosa penyakit terbanyak yaitu terkait diabetes 64 kasus, stroke 64 kasus, dan ginjal 26 kasus, dengan rata rata lama dirawat 1-13 hari dengan kategori pasien umum 91 dan pasien BPJS 486. Pada pengelompokan usia lansia akhir asal kecamatan terbanyak yaitu 83 dari Ngawi, 76 dari Paron. dan 60 dari Geneng.

Pengelompokan usia 65 tahun sampai seterusnya termasuk usia manula, engan jumlah 553, jenis kelamin laki laki 262 dan perempuan 271, diagnosa penyakit terbanyak yaitu terkait stroke 66 kasus, diabetes 37 kasus dan katarak 35 kasus dengan rata rata lama dirawat 1-12 hari dengan kategori pasien umum 78 dan pasien BPJS 475. Pada pengelompokan usia manula asal kecamatan terbanyak yaitu 109 dari Ngawi, 75 dari Paron, dan 53 dari Padas.

### 3.3 Pola Penyebaran Penyakit Pada Kasus Rawat Inap Di Rumah Sakit Widodo Ngawi

Untuk melihat pola persebaran penyakit pada data kunjungan pasien rawat inap di Rumah Sakit Widodo Ngawi, peneliti menganalisis jumlah kunjungan pasien pada asal kecamatan yang ada di kabupaten Ngawi dengan hasil diagnosa kasus terbanyak. Untuk pembuatan pola penyebaran penyakit perlu dilakukan proses pemetaan peta. Proses ini dilakukan dengan ekstraksi data pada peta di Indonesia kemudian diambil peta Kabupaten Ngawi dengan atribut kecamatan. Di kabupaten Ngawi terdapat 19 kecamatan yaitu Karangjati, Kasreman, Geneng, Padas, Paron, Kwadungan, Sine, Jogorogo, Ngawi, Bringin, Kendal, Ngrambe, Karanganyar, Mantingan, Pangkur, Gerih, Kedunggalar, dan Widodaren. Hasil pola penyebaran penyakit dapat dilihat pada Gambar 13 di bawah ini.



Gambar 13. Peta Hasil Pengelompokan Pola Penyebaran Penyakit

## 4. KESIMPULAN

Dari hasil penelitian ini dapat disimpulkan penerapan Data mining melalui *K-Means clustering* pada pengelompokan data rekam medis pasien berdasarkan data variable diagnosa penyakit, kecamatan, usia, jenis kelamin, lama dirawat dan cara bayar dapat diterapkan menggunakan Aplikasi RStudio. Sumber data studi ini ialah data sekunder berdasarkan data pasien rawat inap Bulan Oktober, November, dan Desember tahun 2023 Di Rumah Sakit Widodo Ngawi. Jumlah data yang dipakai 3171 data. Dari hasil pengelompokan diperoleh empat klaster, klaster 1 terdiri dari 524 pasien (19%), klaster 2 terdiri dari 831 pasien (30%), klaster 3 terdiri dari 940 pasien, (33%) dan

klaster 4 terdiri dari 512 pasien (18%). Kasus persebaran penyakit yang paling tinggi diderita pasien rawat inap berdasarkan asal kecamatan yaitu demam 74 kasus, sistem pencernaan 64 kasus, stroke 43 kasus, penyakit degenerative seperti diabetes 64 kasus, jantung 7 kasus, neoplasma 6 kasus, asma 4 kasus dan system sirkulasi 3 kasus, serta penyakit akibat luar seperti cedera dan patah tulang 10 kasus. Pengelompokan data rekam medis pasien dari proses data mining diatas dapat digunakan sebagai dasar manajemen statistika dalam pengambilan keputusan di Rumah Sakit terkait dengan peningkatan pelayanan atau fasilitas pelayanan rawat inap seperti ketersediaan ruang perawatan dan tempat tidur, penyediaan stok obat, ataupun kebutuhan dokter spesialis khususnya terkait penyakit tersebut. Selain itu pengelompokan data rekam medis pasien dari proses data mining dihasilkan pengelompokan penyakit menghasilkan informasi baru mengenai pengelompokan penyebaran penyakit disetiap kecamatan. Hal ini dapat dijadikan acuan bagi pemerintah daerah terutama dinas kesehatan Kabupaten Ngawi untuk melakukan sosialisasi dan pencegahan terhadap sumber penyakit berdasarkan diagnosa penyakit, usia, dan jenis kelamin yang banyak diderita pasien.

#### UCAPAN TERIMA KASIH

Peneliti ucapkan terima kasih kepada STIKES Bhakti Husada Mulia Madiun dan pada pihak-pihak terkait yang turut terlibat dalam penyelesaian studi ini.

#### REFERENSI

- [1] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus : Poli Klinik PT.Inecda)," *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- [2] C. Sanggamele, F. K. Kolibu, and F. R. R. Maramis, "Analisis Pengelolaan Rekam Medis Di Rumah Sakit Umum Pancaran Kasih Manado," *J. KESMAS*, vol. 7, no. 4, pp. 1–11, 2018.
- [3] B. W. Aulia, M. Rizki, P. Prindiyana, and S. Surgana, "Peran Krusial Jaringan Komputer dan Basis Data dalam Era Digital," *JUSTINFO / J. Sist. Inf. dan Teknol. Inf.*, vol. 1, no. 1, pp. 9–20, 2023, doi: 10.33197/justinfo.vol1.iss1.2023.1253.
- [4] D. U. Dedo, M. M. Ngongo, T. Informatika, S. Stella, and M. Sumba, "Promotional Strategy To Increase Computer Sales ( Case Study : Global Teknik ) Kata Kunci : Strategi pemasaran , supply chain Perkembangan teknologi yang berkembang begitu cepat merambah ke berbagai bidang kehidupan manusia . Teknologi informasi tidak dap," vol. 1, no. 2, pp. 131–139, 2023.
- [5] M. Rivki, A. M. Bachtiar, T. Informatika, F. Teknik, and U. K. Indonesia, *Data Mining and Learning Analytic*, no. 112.
- [6] M. Kamber and J. Han, *Data Mining: Concepts and Techniques : Concepts and Techniques*. 2006.
- [7] W. Purba *et al.*, "Penerapan Data Mining Untuk Pengelolaan Data Rekam Medis Menggunakan Metode K-means Clustering Pada Rumah Sakit Royal Prima Medan," *J. TEKINKOM*, vol. 6, no. 1, pp. 158–168, 2023, doi: 10.37600/tekinkom.v6i1.857.
- [8] M. Risqi Ananda, N. Sandra, E. Fadhila, A. Rahma, and N. Nurbaiti, "Data Mining dalam Perusahaan PT Indofood Lubuk Pakam," *Com. Commun. Inf. Technol. J.*, vol. 2, no. 1, pp. 108–119, 2023, doi: 10.47467/comit.v2i1.124.
- [9] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 30–38, 2021, doi: 10.37148/bios.v2i1.19.
- [10] H. Nusa Putra and D. Putri Anisa, "Klasterisasi Data Rekam Medis Pada Diagnosa Penyakit Berdasarkan Usia Pasien Menggunakan Algoritma K-Means Di Puskesmas Lubuk Alung," *Ensiklopedia J.*, vol. 3, no. 5, pp. 128–133, 2021.
- [11] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.
- [12] Yullia Agustin and Adiati Trihastuti, "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Wajib Pajak Kendaraan Bermotor (Di Kantor Samsat Kabupaten Lamongan)," *J. Student Res.*, vol. 1, no. 2, pp. 248–263, 2023, doi: 10.55606/jsr.v1i2.994.
- [13] Yahya and W. P. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape ( Rokok El ektrik ) pada ' Lombok Vape On ' Pendahuluan dihasilkan tidak stabil dan tidak mampu diprediksi Dari penelitian yang dilakukan , berusaha untuk mengklasifikasik," *J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020.
- [14] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022.
- [15] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.