Pengelompokkan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma K-Means (Studi Kasus Puskesmas Parongpong)

Nisa Nurliah[1], Ade Yuliana[2]

[1,2]*Program Studi Teknik Informatika - Politeknik TEDC Bandung*

Jl. Politeknik-Pesantren KM2 Cibabat Cimahi Utara – Cimahi Jawa Barat – Indonesia

adnis1002@gmail.com[1], yulianaad@poltektedc.ac.id[2]

***Abstrak*— Rekam data medis atau biasa di sebut dalam bidang kesehatan adalah ICD (*International Classification Diseases*) merupakan rekaman dari riwayat pasien yang melakukan pengobatan di Rumah Sakit maupun klinik. Puskesmas Parongpong merupakan salah satu fasilitas layanan kesehatan di Kabupaten Bandung Barat yang setiap tahunnya menerima ribuan kunjungan pasien dengan beragam jenis penyakit. Namun, sistem informasi yang tersedia seperti SIMPUS belum mampu menyajikan informasi analitis berupa pengelompokan penyakit yang bisa dimanfaatkan untuk perencanaan layanan kesehatan. Penelitian ini bertujuan untuk mengelompokkan data rekam medis pasien berdasarkan jenis penyakit menggunakan metode clustering dengan algoritma K-Means dan pendekatan CRISP-DM sebagai tahapan analisis data mining. Dataset yang digunakan berjumlah 16.366 data pasien dari Poli Umum Puskesmas Parongpong selama tahun 2023. Data terdiri dari lima atribut utama, yaitu nomor rekam medis, jenis kelamin, usia, wilayah asal pasien, dan nama penyakit. Proses clustering dilakukan dengan aplikasi RapidMiner, dan hasil terbaik diperoleh saat data dikelompokkan menjadi tiga klaster berdasarkan nilai Davies Bouldin Index sebesar-0,826. Hasil penelitian menunjukkan bahwa Klaster 0 didominasi oleh penyakit flu (Acute Nasopharyngitis) yang umum terjadi pada pasien balita, Klaster 1 oleh penyakit Dyspepsia dan hipertensi pada kelompok dewasa dan lansia, serta Klaster 2 oleh penyakit ISPA, diare, dan skabies yang sering menyerang anak-anak dan remaja. Pengelompokan ini memberikan gambaran yang jelas tentang persebaran penyakit berdasarkan usia, jenis kelamin, dan wilayah, serta dapat menjadi landasan pengambilan keputusan dalam peningkatan pelayanan preventif dan kuratif di Puskesmas.**

Kata Kunci*— *Rekam medis, data mining, clustering, K-Means, Puskesmas Parongpong.

Abstract*— *Medical records, also known in the health sector as ICD (International Classification of Diseases), are records of patients' medical histories during treatment in hospitals or clinics. Parongpong Community Health Center is one of the healthcare facilities in West Bandung Regency that receives thousands of patient visits each year with various diseases. However, the available information system, such as SIMPUS, has not been able to provide analytical information in the form of disease groupings that could be utilized for health service planning. This study aims to cluster patient medical record data based on disease types using the K-Means algorithm with the CRISP-DM approach as the stages of data mining analysis. The dataset consisted of 16,366 patient records from the General Clinic of Parongpong Community Health Center in 2023. The data included five main attributes: medical record number, gender, age, region of origin, and disease name. The clustering process was carried out using RapidMiner, and the best results were achieved when the data were grouped into three clusters based on a Davies Bouldin Index value of -0.826. The findings show that Cluster 0 was dominated by influenza (Acute Nasopharyngitis), commonly affecting toddlers; Cluster 1 was dominated by dyspepsia and hypertension, mostly found in adults and the elderly; while Cluster 2 was dominated by acute respiratory infections (ARI), diarrhea, and scabies, which frequently affect children and adolescents. These groupings provide a clear overview of disease distribution by age, gender, and region, and can serve as a basis for decision-making to improve preventive and curative health services at the community health center.

Keywords*— *Medical records, data mining, clustering, K-Means, Parongpong Community Health Center.

I. PENDAHULUAN

Rekam medis merupakan kumpulan informasi yang memuat identitas pasien, pemeriksaan, pengobatan, tindakan medis, serta pelayanan kesehatan lain yang diterima pasien, baik di rumah sakit maupun fasilitas kesehatan tingkat pertama. Dalam praktik internasional, rekam medis dikodekan ke dalam International Classification of Diseases (ICD) untuk memudahkan klasifikasi dan analisis data kesehatan [1].

Puskesmas Parongpong sebagai salah satu fasilitas pelayanan kesehatan di Kabupaten Bandung Barat melayani ribuan kunjungan pasien setiap tahunnya dengan beragam jenis penyakit. Data kunjungan pasien tahun 2023 tercatat mencapai 16.366 pasien pada Poli Umum. Meskipun telah menggunakan Sistem Informasi Puskesmas (SIMPUS), sistem tersebut belum menyediakan analisis pengelompokan penyakit yang dapat dimanfaatkan dalam perencanaan pelayanan kesehatan [2]. Padahal, pengelompokan data penyakit penting untuk mendukung pengambilan keputusan berbasis data, misalnya dalam penyusunan program preventif, promotif, maupun kuratif yang lebih tepat sasaran.

Salah satu pendekatan yang dapat digunakan adalah data mining, khususnya metode clustering. Clustering merupakan teknik pengelompokan data ke dalam beberapa kelompok berdasarkan kemiripan tertentu tanpa adanya label kelas [3].

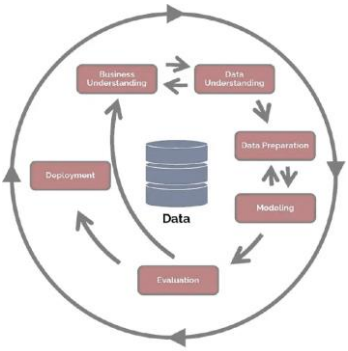
Dari berbagai algoritma clustering, K-Means merupakan salah satu metode yang paling populer karena sederhana, efisien, dan mampu memberikan hasil yang optimal dalam pengelompokan data berskala besar [4].

Beberapa penelitian sebelumnya telah menerapkan K-Means dalam pengelompokan data kesehatan. Wibowo (2023) menunjukkan bahwa algoritma ini mampu membagi data rekam medis pasien di RSUD M. Natsir Kota Solok ke dalam empat klaster berdasarkan jenis penyakit, usia, dan jenis kelamin [5]. Penelitian lain oleh Aria dkk. (2023) juga membuktikan efektivitas K-Means dalam mengelompokkan data penyakit diare di Puskesmas Beber ke dalam tiga klaster utama [6]. Hal ini membuktikan bahwa K-Means layak digunakan untuk analisis data kesehatan di fasilitas pelayanan tingkat pertama, termasuk puskesmas.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengelompokkan data rekam medis pasien di Puskesmas Parongpong berdasarkan jenis penyakit menggunakan algoritma K-Means dengan pendekatan CRISP-DM (Cross-Industry Standard Process for Data Mining). Hasil pengelompokan diharapkan mampu memberikan gambaran pola persebaran penyakit berdasarkan usia, jenis kelamin, dan wilayah, sehingga dapat digunakan sebagai dasar perencanaan peningkatan layanan kesehatan masyarakat di Puskesmas Parongpong.

II. METODE PENELITIAN

Metode penelitian merupakan serangkaian kegiatan dalam mencari kebenaran suatu studi penelitian, yang diawali dengan suatu pemikiran yang membentuk rumusan masalah sehingga menimbulkan hipotesis awal, dengan dibantu dan persepsi penelitian terdahulu, sehingga penelitian bisa diolah dan dianalisis yang akhirnya membentuk suatu kesimpulan [7]. Pada penelitian ini, algoritma k-means digunakan untuk mengelompokkan data rekam medis pasien berdasarkan penyakit. Proses pengelompokan ini di implementasikan dengan pendekatan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang merupakan kerangka kerja standar dalam proyek data mining.



Gbr. 1 Metode CRISP-DM

Penelitian dilaksanakan di Puskesmas Parongpong, Kabupaten Bandung Barat, Jawa Barat, pada bulan Juli 2024. Data yang digunakan adalah data rekam medis pasien Poli Umum selama periode Januari–Desember 2023.

A. *Pengumpulan Data*

Data primer merupakan data yang didapat secara langsung dari sumber data yang berhubungan dengan topik penelitian yang dilaksanakan. Data primer merupakan data yang didapat secara langsung dari sumber data yang berhubungan dengan topik penelitian yang dilaksanakan.

Data sekunder merupakan data yang didapat secara tidak langsung berupa dari sumber lain yang telah ada. Dalam penelitian ini penulis menggunakan data sekunder dari jurnal, tugas akhir terdahulu yang berkaitan dengan pengelompokkan data rekam medis. Algoritma Data Mining, *Association Rules* dengan menggunakan *Algoritma Clustering*.

Dalam penelitian kali ini penulis juga melakukan studi kepustakaan dengan cara mempelajari teori-teori dan buku yang berkaitan dengan implementasi data mining serta selain itu penulis juga mempelajari hasil penelitian Penerapan data mining untk pengelompokkan data rekam medis pasien dengan algoritma clustering.

B. *Dataset*

Dataset yang digunakan berjumlah 16.366 kunjungan pasien pada tahun 2023, dengan lima atribut utama yang dianalisis, yaitu:

1. Nomor rekam medis,
2. Jenis kelamin,
3. Usia,
4. Wilayah asal pasien,
5. Nama penyakit

Atribut lain seperti tanggal kunjungan, jumlah kunjungan per bulan, dan nama pasien tidak digunakan karena bersifat administratif dan tidak relevan terhadap proses clustering.

C. *Preparasi Data*

Data yang sudah diapatkan selanjutnya akan dipreparasi atau cleaning data. Dimana proses ini, data akan diolah kembali untuk menghilangkan data yang tidak konsisten atau menghapus data yang kurang penting pada proses pengolahan data nantinya. *Field* yang digunakan dari data tersebut adalah *field* wilayah, jenis kelamin, umur, diagnosa penyakit. Dimana *field-field* tersebut akan penulis gunakan sebagai indikator analisis untuk hasil penelitian.

TANGGAL	JUMLAH HARIAN	JUMLAH Poli	No. Medrec	Nama Pasien	L/P	Umur	Wilayah/Desa	ICD 10	Nama Penyakit2
02/01/2023	60	umum	05-011939	MUHAMMAD ARYA	L	6 TAHUN 0 BU	CIGUGUR GIRANG	J00	Acute nasopharyngitis [common cold]
03/01/2023	42	umum	03-011935	MUHAMMAD ZAVIE	L	3 TAHUN 6 BU	CIGUGUR GIRANG	J00	Acute nasopharyngitis [common cold]
04/01/2023	54	umum	03-000953	SHANIA PUTRI AZZA	P	5 TAHUN 4 BU	KARYAWANGI	J00	Acute nasopharyngitis [common cold]
05/01/2023	41	umum	02-000866	SASKIA PUTRI RAH	P	16 TAHUN 0 BL	CIHANJUANG RAHAYU	J00	Acute nasopharyngitis [common cold]
06/01/2023	40	umum	02-011943	NOVI HERAWATI, N	P	23 TAHUN 0 BL	KARYAWANGI	M79.1	Myalgia
07/01/2023	17	umum	02-010527	ELANG, Ny	P	56 TAHUN 5 BL	CIHANJUANG RAHAYU	J00	Acute nasopharyngitis [common cold]
09/01/2023	71	umum	05-007941	MUHAMMAD ZAKI	L	2 TAHUN 2 BU	CIHANJUANG RAHAYU	J06.9	Acute upper respiratory infection, unspecified
10/01/2023	38	umum	01-010705	SITI AISAH	P	52 TAHUN 11 B	CIHANJUANG	I10	Essential (primary) hypertension
11/01/2023	46	umum	01-002927	EKI DIRGAYANTI, N	P	48 TAHUN 2 BL	KARYAWANGI	I10	Essential (primary) hypertension
12/01/2023	41	umum	00-000964	KARYA SASMITA	L	71 TAHUN 8 BL	CIHANJUANG RAHAYU	I10	Essential (primary) hypertension
13/01/2023	32	umum	01-011934	YUYUM, NY	P	55 TAHUN 4 BL	KARYAWANGI	I10	Essential (primary) hypertension
14/01/2023	28	umum	03-000577	ZAHIRA KAFAF	P	15 TAHUN 4 BL	CIHANJUANG RAHAYU	J02.9	Acute pharyngitis, unspecified
16/01/2023	55	umum	01-000713	OMI, NY	P	54 TAHUN 5 BL	CIHANJUANG RAHAYU	M79.1	Myalgia
17/01/2023	47	umum	01-000494	IIS ROHAETI,NY	P	34 TAHUN 7 BL	KARYAWANGI	J00	Acute nasopharyngitis [common cold]
18/01/2023	41	umum	01-008121	NETY ROHAENI	P	45 TAHUN 3 BL	CIHANJUANG	K30	Dyspepsia
19/01/2023	32	umum	02-009017	TETI ROHAETI, Ny	P	51 TAHUN 9 BL	CIHANJUANG	M79.1	Myalgia
20/01/2023	37	umum	04-002369	BILAL ALTARIS	L	2 TAHUN 10 BL	CIHANJUANG RAHAYU	J00	Acute nasopharyngitis [common cold]
21/01/2023	30	umum	04-010249	MUHAMMAD ADZR	L	5 TAHUN 2 BU	CIHANJUANG	J00	Acute nasopharyngitis [common cold]
24/01/2023	59	umum	05-008621	TRIAN FIKRIANTO	L	19 TAHUN 3 BL	CIHANJUANG RAHAYU	J06.9	Acute upper respiratory infection, unspecified
25/01/2023	56	umum	01-007145	IPIT KASTINI, Ny	P	38 TAHUN 3 BL	CIHANJUANG RAHAYU	K30	Dyspepsia
26/01/2023	55	umum	01-011068	HARJA	L	54 TAHUN 8 BL	CIHANJUANG	M79.1	Myalgia
27/01/2023	35	umum	01-005758	IDA DARWATI, Ny	P	44 TAHUN 5 BL	KARYAWANGI	A09	Diarrhea and gastroenteritis of presumed infectious origin

Gbr. 2 Data Awal

Pada Gbr. 2 diatas dari 7007 data rekam medis pasien ada 11 atribut yang belum dilakukan preparasi data seperti tanggal kunjungan, jumlah harian pasien, jumlah total harian pasien, Poli, no medrec, nama pasien, jenis kelamin, wilayah/desa. Data rekam medis pasien tersebut masih dalam kondisi awal dan belum melalui proses preparasi, sehingga memerlukan

pembersihan dan pengolahan lebih lanjut sebelum dilakukan analisis clustering.

Untuk menghindari adanya duplikasi, data kosong (missing values), maupun data yang tidak konsisten. Adapun data yang sudah di preparasi dapat dilihat pada Gbr. 3 di bawah ini.

No. Medrec	L/P	Umur	Wilayah/Desa	Nama Penyakit2
05-011939	L	6 TAHUN 0 BULAN	CIGUGUR GIRANG	Acute nasopharyngitis [common cold]
03-011935	L	3 TAHUN 6 BULAN	CIGUGUR GIRANG	Acute nasopharyngitis [common cold]
03-000953	P	5 TAHUN 4 BULAN	KARYAWANGI	Acute nasopharyngitis [common cold]
02-000866	P	16 TAHUN 0 BULAN	CIHANJUANG RAHAYU	Acute nasopharyngitis [common cold]
02-011943	P	23 TAHUN 0 BULAN	KARYAWANGI	Myalgia
02-010527	P	56 TAHUN 5 BULAN	CIHANJUANG RAHAYU	Acute nasopharyngitis [common cold]
05-007941	L	2 TAHUN 2 BULAN	CIHANJUANG RAHAYU	Acute upper respiratory infection, unspecified
01-010705	P	52 TAHUN 11 BULAN	CIHANJUANG	Essential (primary) hypertension
01-002927	P	48 TAHUN 2 BULAN	KARYAWANGI	Essential (primary) hypertension
00-000964	L	71 TAHUN 8 BULAN	CIHANJUANG RAHAYU	Essential (primary) hypertension
01-011934	P	55 TAHUN 4 BULAN	KARYAWANGI	Essential (primary) hypertension
03-000577	P	15 TAHUN 4 BULAN	CIHANJUANG RAHAYU	Acute pharyngitis, unspecified
01-000713	P	54 TAHUN 5 BULAN	CIHANJUANG RAHAYU	Myalgia
01-000494	P	34 TAHUN 7 BULAN	KARYAWANGI	Acute nasopharyngitis [common cold]
01-008121	P	45 TAHUN 3 BULAN	CIHANJUANG	Dyspepsia
02-009017	P	51 TAHUN 8 BULAN	CIHANJUANG	Myalgia
04-002369	L	2 TAHUN 10 BULAN	CIHANJUANG RAHAYU	Acute nasopharyngitis [common cold]
04-010249	L	5 TAHUN 2 BULAN	CIHANJUANG	Acute nasopharyngitis [common cold]
05-008621	L	19 TAHUN 3 BULAN	CIHANJUANG RAHAYU	Acute upper respiratory infection, unspecified
01-007145	P	38 TAHUN 3 BULAN	CIHANJUANG RAHAYU	Dyspepsia
01-011068	L	54 TAHUN 8 BULAN	CIHANJUANG	Myalgia
01-005758	P	44 TAHUN 5 BULAN	KARYAWANGI	Diarrhea and gastroenteritis of presumed infectious origin

Gbr. 3 Data Yang Sudah Di Preparasi

Setelah melakukan preparasi data dari 11 atribut menjadi 5 atribut karena dalam penelitian ini hanya membutuhkan 5 atribut untuk pengolahan data yaitu no medrec, jenis kelamin, umur, wilayah/desa dan nama penyakit.

Agar data rekam medis tersebut dapat diolah menggunakan K- means clustering, maka data yang berjenis non-numerik seperti wilayah, jenis kelamin, umur, diagnosa penyakit, harus di transformasikan terlebih dahulu kedalam bentuk numeric atau angka yang dimulai dari 0. Ada 4 wilayah yang ada di kelurahan Parongpong yang akan diteliti oleh penulis.

TABEL I
KATEGORI WILAYAH DAN TRANFORMASI

Wilayah	Tranformasi
Desa Cihanjuang	1
Desa Karyawangi	2
Desa Cihanjuang Rahayu	3
Desa Cihideung	4

Mengubah data kategorikal menjadi numerik agar dapat diproses oleh algoritma K-Means. Contohnya: jenis kelamin dikodekan menjadi 0 untuk perempuan dan 1 untuk laki-laki, kategori usia dikodekan sesuai rentang umur (balita, anak-anak, remaja, dewasa, lansia, dan manula), serta wilayah dikodekan berdasarkan desa asal pasien.

D. Hasil

Untuk menilai kualitas hasil clustering, digunakan metrik evaluasi Davies Bouldin Index (DBI). Nilai DBI menunjukkan seberapa baik suatu cluster terbentuk, dengan kriteria semakin kecil nilai DBI maka semakin baik kualitas clustering. Pada penelitian ini, berbagai nilai K dicoba, kemudian dipilih jumlah cluster dengan DBI terbaik sebagai hasil akhir.

Dataset rekam medis pasien berdasarkan Penyakit dikelompokkan berdasarkan atribut, data dan atribut tersebut dikelolah menggunakan aplikasi RapidMiner dengan menggunakan algoritma Clustering K-Means yang akan menghasilkan cluster. Hasilnya akan mengelompokkan data sesuai dengan kemiripan data.

E. Analisis

Hasil clustering ini dapat digunakan sebagai dasar bagi Puskesmas Parongpong untuk merancang kebijakan kesehatan yang lebih terfokus, baik dalam aspek pelayanan kuratif maupun preventif. Puskesmas dapat menggunakan hasil pengelompokan ini untuk memahami tren penyakit, memetakan wilayah yang rentan, serta merencanakan tindakan medis yang lebih terarah sesuai kebutuhan spesifik populasi.

III. HASIL DAN PEMBAHASAN

Data rekam medis pasien berdasarkan jenis penyakit yang sudah di preparasi dapat dilihat pada Gbr. 4 di bawah ini.

L/P	Umur	Wilayah/Desa	Nama Penyakit2
L	6 TAHUN 0 BULAN	CIGUGUR GIRANG	Acute nasopharyngitis [common cold]
L	3 TAHUN 6 BULAN	CIGUGUR GIRANG	Acute nasopharyngitis [common cold]
P	5 TAHUN 4 BULAN	KARYAWANGI	Acute nasopharyngitis [common cold]
P	16 TAHUN 0 BULAN	CIHANJUANG RAHAYU	Acute nasopharyngitis [common cold]
P	23 TAHUN 0 BULAN	KARYAWANGI	Myalgia
P	56 TAHUN 5 BULAN	CIHANJUANG RAHAYU	Acute nasopharyngitis [common cold]
L	2 TAHUN 2 BULAN	CIHANJUANG RAHAYU	Acute upper respiratory infection, unspecified
P	52 TAHUN 11 BULAN	CIHANJUANG	Essential (primary) hypertension
P	48 TAHUN 2 BULAN	KARYAWANGI	Essential (primary) hypertension
L	71 TAHUN 8 BULAN	CIHANJUANG RAHAYU	Essential (primary) hypertension
P	55 TAHUN 4 BULAN	KARYAWANGI	Essential (primary) hypertension
P	15 TAHUN 4 BULAN	CIHANJUANG RAHAYU	Acute pharyngitis, unspecified
P	54 TAHUN 5 BULAN	CIHANJUANG RAHAYU	Myalgia
P	34 TAHUN 7 BULAN	KARYAWANGI	Acute nasopharyngitis [common cold]
P	45 TAHUN 3 BULAN	CIHANJUANG	Dyspepsia
P	51 TAHUN 8 BULAN	CIHANJUANG	Myalgia
L	2 TAHUN 10 BULAN	CIHANJUANG RAHAYU	Acute nasopharyngitis [common cold]
L	5 TAHUN 2 BULAN	CIHANJUANG	Acute nasopharyngitis [common cold]
L	19 TAHUN 3 BULAN	CIHANJUANG RAHAYU	Acute upper respiratory infection, unspecified
P	38 TAHUN 3 BULAN	CIHANJUANG RAHAYU	Dyspepsia
L	54 TAHUN 8 BULAN	CIHANJUANG	Myalgia
P	44 TAHUN 5 BULAN	KARYAWANGI	Diarrhea and gastroenteritis of presumed infectious origin

Gbr. 4 Data Preparasi Rekam Medis Pasien Berdasarkan Jenis Penyakit

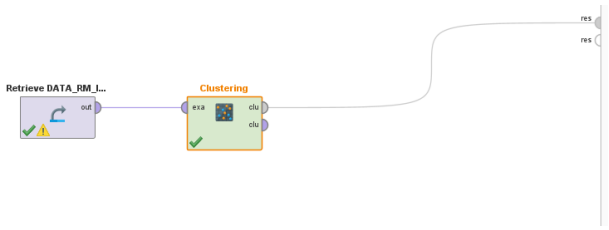
Sebagaimana dapat dilihat pada Gambar 4.2 dan 4.3 diatas, dibuat inisialisasi sebagai berikut:

TABEL II
KATEGORI INISIALISASI

No.	Atribut	Inisialiasai	Keterangan
1.	Nama pasien	np1-np600	Nama pasien
2.	Umur	Balita=1 Kanak-Kanak=2 Remaja Awal=3 Remaja Akhir=4 Dewasa Awal=5 Dewasa Akhir=6 Lansia Awal=7 Lansia Akhir =8 Manula=9	Klasifikasi Umur
3.	L/P	Perempuan=0 Laki-laki=1	Jenis Kelamin
4.	Wilayah/desa	Desa Cihanjuang=1 Desa Cihanjuang Rahayu=2 Desa Cihideung=3 Desa Karyawangi=4	Klasifikasi Wilayah/desa
5.	Diagnosa/nama penyakit	Acute nasopharyngitis [common cold]=1 Essential (primary) hypertension=2 Dyspepsia=3 Acute upper respiratory infection, unspecified=4 Myalgia=5	Klasifikasi penyakit

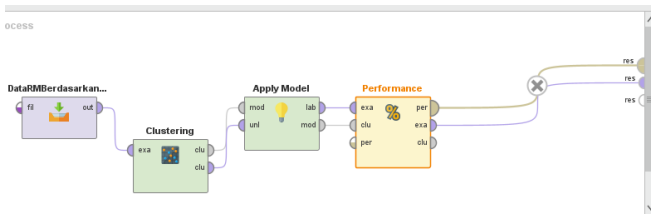
No.	Atribut	Inisialiasai	Keterangan
		Diarrhea and gastroenteritis of presumed infectious origin=6 Scabies=7 Dermatititis=8 Acute pharyngitis, unspecified=9 Fever, unspecified=10	

Proses pengujian ini menggunakan *software Rapidminer* dan mengaplikasikan Algoritma K-Means dengan memasukan data yang sudah di preparasi sebelumnya. Proses pengujian tersebut tersebut dapat di lihat pada Gbr 5. berikut.



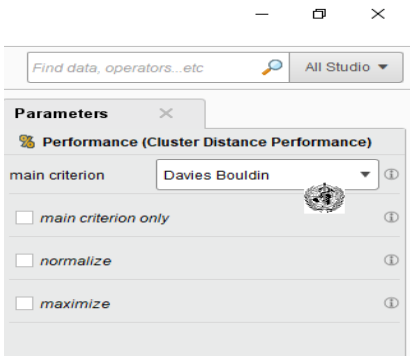
Gbr. 5 Proses Memasukan Dataset Ke Dalam RapidMinner

Kemudian untuk melanjutkan proses selanjutnya maka di perlukan penentuan kluster terbaik dengan cara mencari nilai davis bouldin sehingga nanti akan di temukan nilai k terbaik dengan menambahkan Performance pada proses rapidminer. Serta menggunakan Apply model untuk menghasilkan Kesimpulan tentang hubungan Proses tersebut dapat di lihat pada Gbr. 6 di bawah ini.



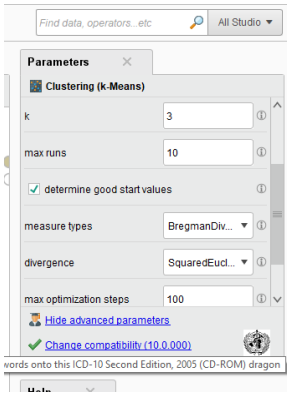
Gbr. 6 Menambahkan *Performance* Dan *Apply Model*

Setelah di tambahkan performance untuk mendapatkan hasil yang di inginkan maka perlu di atur beberapa parameter nya. Dikarenakan pada proses ini hasil yang di inginkan oleh peneliti adalah nilai davies bouldin maka pada parameter performance di ubah menjadi Davies Bouldin seperti Gbr. 7 di bawah ini.



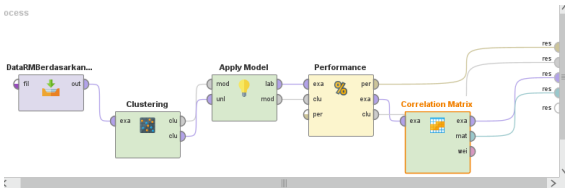
Gbr. 7 Proses Mengatur *Parameters* Pada *Perpormance Cluster*

Setelah beberapa proses di atas terlewati maka selanjutnya adalah mengatur parameter pada K-means clustering untuk mencari nilai kluster terbaik. Penulis melakukan proses dengan menggunakan metode eksperimen dimana peneliti akan memasukan beberapa kluster secara acak sehingga nantinya akan di dapatkan nilai kluster mana yang terbaik. Proses pengaturan parameter k-means klustering nya dapat di lihat pada Gbr. 8 di bawah ini.



Gbr. 8 Proses Mengatur *Parameters Clustering K-Means*

Tahap akhir yang di perlukan untuk penentuan kluster terbaik dengan cara mencari nilai Corelation matrix sehingga nanti akan di temukan nilai k terbaik dengan menambahkan Corelation matrix pada proses rapidminer. Proses tersebut dapat di lihat pada Gbr. 9 di berikut ini.



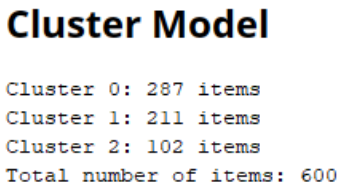
Gbr. 9 Menambahkan *Corealation Matrix*

Dari proses penentuan kluster di dapatkan kluster dua dengan nilai Davies Bouldin Index (DBI) terbaik. Maka pengelompokan data pada penelitian kali ini di bagi menjadi 4 (empat) cluster yaitu :

1. Kluster 0 dengan jumlah data 287 data
2. Kluster 1 berjumlah 211

3. Kluster 2 berjumlah 102
4. Kluster 3 berjumlah 201

Namun setelah di uji oleh aplikasi RapidMiner hasil uji terbaik ada pada kluster 3 (tiga). Pengelompokan data tersebut dapat di lihat pada Gbr. 10 di bawah ini.



Gbr. 10 Pengelompokan Data Sesuai *Cluster*

Berikut hasil data rekam medis yang termasuk kedalam *cluster* 0, *cluster* 1, *cluster* 2 dapat di lihat pada TABEL III di bawah ini.

TABEL III
TABEL HASIL *CLUSTER*

Cluster_0	
Jenis Penyakit	Umur
Acute nasopharyngitis [common cold],	Balita=92 Remaja Awal=71 Remaja Akhir=86 Kanak-kanak=38

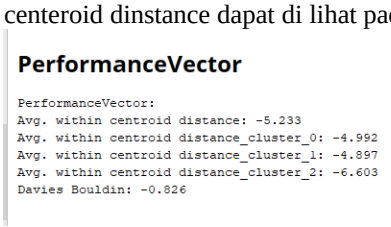
Cluster_1	
Jenis Penyakit	Umur
Dyspepsia	Balita=0 Lansia Awal=59 Lansia Akhir=42 Dewasa Akhir=52 Manula=27
Essential(primary) hypertension	Balita=0 Lansia Awal =92 Lansia Akhir =56 Dewasa Akhir =51 Manula =45

Cluster_2	
Acute upper respiratory infection, unspecified	Remaja Akhir=9 Balita=7 Kanak-kanak=10
Diarrhea and gastroenteritis of presumed infectious origin	Remaja Akhir=11 Kanak-kanak=12 Manula=14 Lansia Awal=5
Scabies	Remaja Akhir=12 Balita=0 Kanak-kanak=10 Remaja Awal=11
Myalgia	Balita=0 Kanak-kanak=0 Dewasa Akhir=6 Lansia Akhir=6

Adapun hasil pengelompokan adalah sebagai berikut:

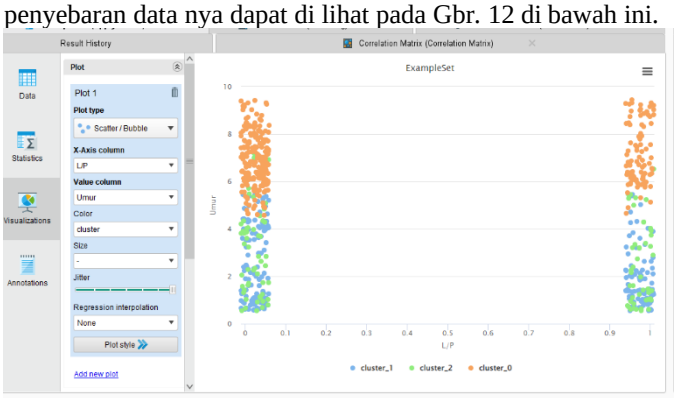
- **Klaster 0**: Didominasi oleh penyakit *Acute Nasopharyngitis* (flu/batuk pilek), terutama menyerang pasien berusia balita.
- **Klaster 1**: Didominasi oleh *Dyspepsia* (gangguan lambung) dan *Hipertensi*, yang lebih banyak terjadi pada pasien dewasa dan lansia.
- **Klaster 2**: Didominasi oleh penyakit *Infeksi Saluran Pernapasan Akut (ISPA)*, diare/gastroenteritis, serta *Scabies*, yang paling sering menyerang anak-anak dan remaja.

Dimana untuk akurasi berdasarkan rata-rata within centeroid dinstance dapat di lihat pada Gbr. 11.



Gbr. 11 Hasil *PerformaceVector* Dengan Nilai DBI

Sedangkan untuk grapik Plot Data View berdasarkan cluster yang telah di tentukan di awal yaitu berjumlah 3 cluster dimana ada cluster 0, cluster 1 dan cluster 2 maka grafik penyebaran data nya dapat di lihat pada Gbr. 12 di bawah ini.



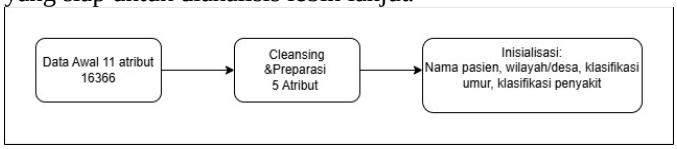
Gbr. 12 Plot Data Penyebaran Data Berdasarkan *Cluster*

Hasil clustering memperlihatkan bahwa terdapat perbedaan distribusi penyakit berdasarkan usia, jenis kelamin, dan wilayah. Misalnya, penyakit flu lebih sering ditemukan pada kelompok usia balita, sedangkan hipertensi banyak dialami oleh kelompok usia lanjut. Hal ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa faktor usia dan lingkungan berpengaruh signifikan terhadap pola penyakit [1].

Selain itu, wilayah dengan kepadatan penduduk tinggi cenderung memiliki kasus ISPA lebih banyak, sedangkan penyakit pencernaan seperti diare banyak ditemukan pada

wilayah dengan kondisi sanitasi kurang baik. Pola ini memberikan gambaran nyata bahwa faktor demografi dan lingkungan merupakan variabel penting dalam analisis kesehatan masyarakat. Dengan mengikuti metode CRISP-DM, proses analisis ini tidak hanya menghasilkan informasi yang berharga tentang pola penyakit di Puskesmas Parongpong tetapi juga memberikan dasar yang kuat untuk pengambilan keputusan dalam kebijakan kesehatan masyarakat.

Dalam penelitian ini, data rekam medis pasien yang digunakan berasal dari Puskesmas Parongpong. Pada tahap awal, jumlah data rekam medis yang diperoleh adalah 16,366 data dengan 11 atribut. Setelah dilakukan proses data cleansing untuk menyaring atribut yang tidak relevan dengan tujuan penelitian, data disederhanakan menjadi 5 atribut utama: nama pasien, jenis kelamin (L/P), umur, wilayah (desa), dan jenis penyakit. Hasil akhir dari proses ini menghasilkan 600 data yang siap untuk dianalisis lebih lanjut.



Gbr. 13 Proses Preparasi Data Rekam Medis pasien berdasrkan jenis penyakit

Hasil clustering dapat dimanfaatkan oleh Puskesmas Parongpong untuk:

1. **Perencanaan program preventif**, misalnya penyuluhan kesehatan lingkungan untuk mencegah ISPA dan diare pada anak-anak.
2. **Layanan kuratif yang lebih terarah**, seperti pemeriksaan tekanan darah rutin bagi kelompok usia dewasa dan lansia untuk mencegah hipertensi.
3. **Manajemen data kesehatan**, yang lebih terstruktur sehingga memudahkan evaluasi kinerja layanan kesehatan dan pembuatan kebijakan berbasis data.

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian serta hasil pembahasan yang di lakukan oleh peneliti tentang pengelompokan data rekam medis pasien berdasarkan jenis penyakit menggunakan K-Means clustering, maka dapat di simpulkan beberapa hal sebagai berikut:

1) Berdasarkan hasil pengujian dengan algoritma K-means, Algoritma K-Means berhasil mengelompokkan data rekam medis pasien menjadi tiga cluster utama berdasarkan kesamaan karakteristik pasien dan jenis penyakitnya. Cluster 0 didominasi oleh penyakit Acute Nasopharyngitis (common cold/flu) yang paling banyak terjadi pada kelompok usia balita. Cluster 1 banyak berisi pasien dengan penyakit Dyspepsia (gangguan lambung/maag) dan Essential Hypertension (hipertensi), yang umumnya diderita oleh dewasa akhir dan lansia. Cluster 2 mengelompokkan pasien dengan penyakit ISPA (Infeksi Saluran Pernapasan Akut), diare/gastroenteritis, dan scabies, yang paling sering terjadi pada anak-anak dan remaja. Dari hasil ini dapat diketahui bahwa penyakit yang paling menonjol di wilayah kerja Puskesmas Parongpong adalah penyakit pernapasan (ISPA dan flu), hipertensi, gangguan pencernaan (dyspepsia dan diare), serta penyakit kulit (scabies dan dermatitis).
2) Berdasarkan pengujian algoritma K-Means berdasarkan Davies Bouldin Index (DBI) menunjukkan bahwa jumlah cluster terbaik adalah K=3 dengan nilai DBI -0,826. Nilai ini menjadi indikasi bahwa kualitas cluster yang terbentuk cukup baik, dengan pemisahan antar cluster yang jelas. pengujian dilakukan dengan mencoba beberapa nilai cluster (K). Hasil evaluasi menggunakan Davies Bouldin Index (DBI) menunjukkan bahwa jumlah cluster K=3 memberikan nilai DBI paling rendah dibandingkan konfigurasi lainnya, sehingga dapat disimpulkan bahwa 3 cluster merupakan jumlah yang paling optimal untuk dataset rekam medis ini. Hasil clustering ini memberikan gambaran pola penyakit berdasarkan kategori usia, jenis kelamin, dan wilayah pasien. Informasi ini dapat dimanfaatkan oleh Puskesmas Parongpong untuk menentukan prioritas penanganan penyakit yang paling dominan, menyusun program promotif dan preventif yang lebih terarah, misalnya penyuluhan kesehatan lingkungan untuk mencegah ISPA dan diare pada anak-anak, serta pemeriksaan rutin tekanan darah pada kelompok dewasa dan lansia.

Pada penelitian kali ini mungkin terdapat banyak kekurangan dalam pelaporan maupun hasil pengujian pada penelitian kali ini. Maka dari itu untuk pengembangan lebih lanjut maka di perlukan beberapa saran sebagai berikut:

1) Untuk pihak Poli umum Puskesmas Parongpong diharapkan agar hasil penelitian ini dapat dikembangkan lagi baik dengan mengimplementasikan kedalam sebuah program agar lebih memudahkan bagian pelayanan Poli umum Puskesmas Parongpong.
2) Untuk pengembangan aplikasi data mining selanjutnya dapat menggunakan algoritma lainnya agar dapat dijadikan perbandingan dari penelitian sebelumnya.

V. DAFTAR PUSTAKA

[1] R. Ordila & R. Wahyuni, “Penerapan Data Mining untuk Pengelompokan Data Rekam Medis Pasien Menggunakan Algoritma K-Means,” Jurnal Rekam Medis dan Informasi Kesehatan, vol. 12, no. 2, pp. 45–52, 2020.

[2] Kementerian Kesehatan RI, “Sistem Informasi Puskesmas (SIMPUS),” Jakarta, 2019.

[3] D. T. Larose & C. D. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, 2nd ed. Hoboken, NJ: Wiley, 2014.

[4] J. Han, M. Kamber, & J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA: Morgan Kaufmann, 2012.

[5] T. Wibowo, “Penerapan Data Mining untuk Pengelompokkan Data Rekam Medis Pasien dengan Algoritma K-Means (Studi Kasus RSUD M. Natsir Kota Solok),” Skripsi, Universitas Negeri Padang, 2023.

[6] A. Aria, F. Salsabila, & D. R. Putri, “Clustering Penyakit Diare Menggunakan Algoritma K-Means (Studi Kasus: Puskesmas Beber),”

Jurnal Teknologi dan Sistem Informasi Kesehatan, vol. 5, no. 1, pp. 21–28, 2023.
[7] Sahir, S., "METODOLOGI PENELTIAN (M. s. Dr.Ir. Koryati (ed.); Cetakan I)". KBM Indonesia, 2021.
[8] Moeloek, N. F. (2019). Permenkes nomor 43 tahun 2019 tentang puskesmas.
[9] Dennis, A. (2013). Belajar Data Mining dengan RapidMiner. In R. Sanjaya (Ed.), Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2 (Vol. 5, Issue 4).
[10] Rahayu, N. S. (2021). Data Mining Dan Penerapan Algoritma (R. W. dan J. Simarmata (ed.); Cetakan 1).
[11] Ali, A. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode KMeans Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo.
MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer, 19(1),186–195. https://doi.org/10.30812/matrik.v19i1.529.