

## Klasterisasi Pasien Rawat Jalan di Puskesmas dengan Menggunakan Metode Algoritma Clustering K-Means

Azkar\*, Kusrini

Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

Email: <sup>1</sup>\*azkars2pjti@students.amikom.ac.id, <sup>2</sup>kusrini@amikom.ac.id

Email Penulis Korespondensi: azkars2pjti@students.amikom.ac.id

**Abstrak**—Puskesmas Aikmel Utara Kabupaten Lombok Timur mulai beroperasi pertengahan tahun 2019, belum pernah memanfaatkan data pasien untuk memperoleh informasi sebagai bahan/dasar dalam pengambilan keputusan dalam upaya peningkatan mutu pelayanan kesehatan sehingga dapat meningkatkan kepuasan pasien. Adapun yang menjadi sasaran atau responden survei kepuasan pasien di Puskesmas adalah pengunjung yang sebagian besar pengunjung rawat jalan. Tujuan penelitian ini melakukan pengelompokan pasien rawat jalan berdasarkan variabel jenis kelamin, umur, status kepesertaan BPJS kesehatan dan alamat pasien serta diagnosa penyakit pasien dengan menggunakan metode algoritma clustering k-means dan bantuan aplikasi *RapidMiner*. Data pasien sebanyak 1889 data dikelompokkan menjadi 2 cluster, 3 cluster, 4 cluster serta 5 cluster, dan dilakukan evaluasi dengan *Davies-Bouldin Indeks* (DBI). Hasil penelitian menunjukkan bahwa jumlah klaster yang dibentuk sebanyak 2 klaster, yaitu klaster 1 sebanyak 1570 data dan klaster 2 sebanyak 319 data. Klaster 1 didominasi oleh pasien perempuan 1074 (68,4%), pasien peserta BPJS 819 (52,2%), dengan kelompok usia terbanyak usia dewasa 883 (56,2%), dan terbanyak dari desa Toya 488 (31,1%), serta diagnosa penyakit terbanyak ISPA akut (J06) 223 (14,2%). Sedangkan klaster 2 didominasi oleh pasien perempuan 205 (64,3%), pasien peserta BPJS 202 (63,3%), dengan kelompok usia terbanyak usia dewasa 191 (59,9%), dan terbanyak dari desa Toya 108 (33,9%), serta diagnosa penyakit terbanyak pemeriksaan kehamilan (Z34) 29 (9,1%). Dari hasil klaster ini diperoleh informasi bahwa pengunjung rawat jalan Puskesmas Aikmel Utara terbanyak adalah dari desa Toya dan didominasi kelompok usia dewasa dan peserta BPJS kesehatan, serta penyakit terbanyak adalah ISPA akut, sehingga diharapkan dapat membantu pihak Puskesmas dalam pengambilan keputusan atau kebijakan yang berkaitan dengan program-program kesehatan di wilayah kerjanya.

**Kata Kunci:** Data mining; Algoritma; Clustering k-means; RapidMiner; Pasien rawat jalan

**Abstract**—The Aikmel Utara Public Health Center in Lombok Timur Regency began operating in mid-2019, but has never utilized patient data to obtain information as a basis for decision-making efforts to improve quality of health services and consequently enhance patient satisfaction. The target respondents patient satisfaction survey at the Public Health Center are visitors, the majority of whom are outpatient visitors. The purpose of this research is to group outpatient patients based on variables such as gender, age, participation status in the BPJS health insurance program and patient address, as well as diagnosis of the patient's disease using the k-means clustering algorithm method with the assistance of the RapidMiner application. Patient data totaling 1889 were grouped into 2 clusters, 3 clusters, 4 clusters, and 5 clusters, and evaluated using the Davies-Bouldin Index (DBI). The research results show that the number of clusters formed is 2, with cluster 1 consisting of 1570 data and cluster 2 consisting of 319 data. Cluster 1 is dominated by female patients (1074 or 68.4%), BPJS participants (819 or 52.2%), with the most common age group being adults (883 or 56.2%), and most of them are from Toya village (488 or 31.1%), with the most common diagnosis being acute respiratory infections (J06) (223 or 14.2%). Meanwhile, cluster 2 is dominated by female patients (205 or 64.3%), BPJS participants (202 or 63.3%), with the most common age group being adults (191 or 59.9%), and most of them are from Toya village (108 or 33.9%), with the most common diagnosis being pregnancy examinations (Z34) (29 or 9.1%). From these cluster results, it is obtained that the majority of outpatient visitors at the Aikmel Utara Public Health Center are from Toya village and are dominated by the adult age group and BPJS health insurance participants, with the most common disease being acute respiratory infections. It is hoped that this information can assist the Public Health Center in making decisions or policies related to health programs in its working area.

**Keywords:** Data Mining; Algorithm; Clustering K-Means; Rapidminer; Outpatient Patients

### 1. PENDAHULUAN

Berdasarkan Permenkes nomor 31 tahun 2019 tentang Sistem Informasi Puskesmas dan Permenkes nomor 43 tahun 2019 tentang Puskesmas, bahwa setiap Puskesmas harus menyelenggarakan sistem informasi Puskesmas. Sistem Informasi Puskesmas merupakan suatu tatanan yang menyediakan informasi untuk membantu proses pengambilan keputusan dalam melaksanakan manajemen Puskesmas dalam upaya mencapai sasaran kegiatannya[1]. Salah satu sumber data yang tersedia di Puskesmas salah satunya adalah data pengunjung/pasien yang dicatat dalam bentuk rekam medis. Dalam rekam medis terdapat catatan tentang identitas dan penyakit pasien[2][3].

Puskesmas Aikmel Utara merupakan salah satu dari 35 Puskesmas yang ada di Kabupaten Lombok Timur, yang mulai beroperasi pada pertengahan tahun 2019, yang dalam pelaksanaan sistem pencatatan dan pelaporannya masih banyak menggunakan sistem pencatatan secara manual, termasuk rekam medis. Sementara jumlah kunjungan/pasien di Puskesmas cukup tinggi yaitu rata-rata 250 pasien per bulan, adapun jumlah kunjungan Puskesmas Aikmel Utara selama satu tahun (tahun 2023) sebanyak 3.503 kunjungan, dengan sepuluh jenis penyakit terbanyak ISPA (J06), DM (E11), Hipertensi (I10), Gastritis (K29.1), Diare/GE (A09), Scabies (B86), Mual muntah (R11), Faringitis (J02), Nyeri sendi (M25) dan Migraen (A01)[4][5]. Puskesmas Aikmel Utara belum pernah memanfaatkan data pengunjung/pasien untuk memperoleh informasi sebagai bahan/dasar dalam pengambilan keputusan dalam upaya peningkatan mutu pelayanan sehingga dapat meningkatkan kepuasan pasien yang menjadi salah satu indikator nasional mutu Puskesmas. Adapun yang menjadi sasaran atau responden kepuasan pasien di Puskesmas adalah pengunjung Puskesmas dan sebagian besar pengunjung Puskesmas adalah rawat jalan.

Dengan melakukan pengolahan data pengunjung/pasien rawat jalan, bisa didapatkan informasi misalnya informasi tentang pola pengunjung/pasien atau pola penyakit pasien. Sementara dalam pengambilan suatu keputusan harus berdasarkan data dan informasi, sehingga keputusan yang diambil tepat sasaran. Pengambil keputusan sering menuntut supaya penyusunan rekomendasi kebijakan dilaksanakan secara cepat dan tepat yang didukung oleh data dan/atau informasi yang akurat sehingga rekomendasi yang diusulkan atau direncanakan benar-benar sesuai dengan situasi dan kondisi yang ada di lapangan. Data mining merupakan proses pencarian pola ataupun informasi menarik dalam suatu data dengan menggunakan teknik atau metode tertentu. Adapun teknik, metode ataupun algoritma dalam data mining itu bervariasi. Dalam pemilihan metode atau algoritma yang digunakan tergantung dari tujuan serta proses *knowledge discovery in databases* (KDD) secara menyeluruh[6]. Data mining bertujuan untuk mencari atau menemukan suatu pola ataupun informasi yang sebelumnya tidak diketahui. Apabila pola-pola tersebut sudah didapatkan maka dapat digunakan untuk menyelesaikan berbagai permasalahan yang ada. Pada umumnya, data mining berkaitan dengan bagaimana cara penyelesaian masalah dengan melakukan analisis data yang ada pada database[7][8]. Berdasarkan tugas yang dapat dilakukan, data mining dikelompokkan menjadi beberapa jenis yakni: deskripsi, estimasi, prediksi, klasifikasi, klasterisasi, asosiasi. Proses pengolahan data pada data mining membutuhkan algoritma-algoritma dalam melakukan ekstraksi sehingga diperoleh informasi atau pola ataupun pengetahuan. Penggunaan algoritma dalam data mining diklasifikasi berdasarkan perannya masing-masing. Beberapa algoritma data mining yang umum digunakan diantaranya algoritma pengelompokan atau *clustering*. Beberapa jenis clustering diantaranya *K-Means Clustering*, *Density-Based Clustering*, *Partitioning Around Medoids* (PAM) Clustering. Dari berbagai jenis clustering yang ada, yang paling populer adalah *k-means*[9]. *K-means clustering* merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam cluster yang lain[10]. *Clustering k-means* banyak digunakan dalam bidang pelayanan kesehatan. Rimelda, dkk, (2021) dalam “Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia” bahwa *K-Means* lebih unggul dibandingkan dengan DBSCAN dalam pengelompokan kasus Covid-19. Algoritma K-Means memiliki nilai terbaik sebesar 0,6902 yang terdapat pada percobaan dengan nilai  $k = 8$ [11]. Zubair, dkk, (2022) dalam “An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling”, bahwa metode/algoritma clustering k-means dapat mengurangi jumlah literasi secara signifikan dengan menentukan sentroid secara sistematis untuk menghasilkan klaster[12].

Algoritma *K-Means* adalah salah satu algoritma *clustering* yang umum digunakan untuk mengelompokkan data berdasarkan kesamaan/kemiripan karakteristik atau ciri-ciri yang serupa/sejenis. Algoritma k-means juga tidak terlalu sulit dalam pengimplementasiannya serta mempunyai kompleksitas waktu dan ruang yang relatif sedikit. Algoritma ini juga merupakan algoritma yang cukup efisien dalam komputasinya serta memberikan hasil yang cukup baik dan memuaskan jika klasternya *compact*, *hyperspherical* dalam *shape* serta bisa memisahkan fitur-fitur ruangnya secara baik[13][14]. Sehingga dengan melakukan *clustering*, Puskesmas dapat mengidentifikasi pola-pola dalam data pengunjung sehingga dapat menyediakan layanan yang lebih tepat dan spesifik untuk setiap kelompok pasien. Melalui pengelompokan ini, Puskesmas bisa memahami jenis layanan kesehatan yang paling dibutuhkan oleh setiap kelompok pasien. Dengan mengetahui kelompok pengunjung/pasien yang berbeda, Puskesmas dapat mengalokasikan sumber daya dengan lebih efisien. Misalnya, jika ditemukan kelompok pasien tertentu yang memiliki kebutuhan layanan yang sama, Puskesmas dapat merancang program-program khusus yang sesuai untuk kelompok tersebut. Maka penelitian ini bertujuan untuk melakukan pengelompokan pengunjung/pasien rawat jalan di Puskesmas berdasarkan jenis kelamin, kelompok umur, status kepesertaan BPJS kesehatan, dan alamat pasien, serta diagnosa jenis penyakit pasien..

Dalam penelitian ini penulis akan melakukan pengelompokan pasien rawat jalan dengan menggunakan algoritma *clustering k-means* di Puskesmas Aikmel Utara Kabupaten Lombok Timur. Pada penelitian-penelitian terdahulu sudah ada yang membahas tentang pengelompokan data pengunjung pasien atau data lainnya dengan menggunakan Algoritma *Clustering (Clustering K-Means)* atau algoritma lainnya. Menurut penelitian Amir Ali & Lilis Masyfufah (2021) yang berjudul “Klasterisasi Pasien BPJS dengan Metode K-Means Clustering Guna Menunjang Program Jaminan Kesehatan Nasional di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo”, bertujuan untuk mendapatkan pola atau informasi baru dari data rekam medis pasien rawat inap peserta BPJS. Data diperoleh dengan mengelompokkan data dengan menggunakan metode/teknik data mining algoritma k-means. Penelitiannya memperoleh 3 klaster: klaster 1 didapatkan pasien perempuan sebanyak 91 orang (51%), klaster 2 didapatkan pasien perempuan sebanyak 26 orang (14%), serta klaster 3 didapatkan pasien laki-laki sebanyak 63 orang (35%). Diagnosis penyakit yang banyak ditemukan yaitu Born in Hospital, Diarrhoea, dan Hemorrhagic. Peneliti menyimpulkan bahwa diagnose penyakit yang banyak diderita oleh pasien peserta BPJS adalah penyakit dengan kode ICD X Z38.0, A09+E86, dan A91, sehingga dapat dimanfaatkan dalam perencanaan menentukan anggaran rumah sakit. Peneliti juga menyarankan untuk peneliti selanjutnya bisa menggunakan kolaborasi dengan metode clustering hierarki supaya menghasilkan hasil pengelompokan data yang lebih baik[15].

Selanjutnya Novi Karolina pada tahun 2021 dalam penelitian yang berjudul “Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering”, menemukan suatu sistem yang akan dipakai untuk mengelompokkan pasien rawat inap menurut kelas kepesertaan BPJS dan variabel lainnya yang ditentukan dengan teknik atau metode *clustering*. Dari hasil uji dengan metode algoritma *clustering k-means*, didapatkan 3 klaster, dimana kelompok yang mempunyai nilai yang paling tinggi data pasien kelas BPJS adalah klaster 1 sebanyak 435 data pasien peserta BPJS pada usia 20-39 tahun dengan jenis penyakitnya tumor jinak, serta kelas BPJS yang digunakan BPJS tingkat 2. Kelompok yang mempunyai nilai menengah data pasien kelas BPJS adalah klaster 2 sebanyak 395 data pasien rawat

inap peserta BPJS pada usia 20-39 tahun dengan jenis penyakit komplikasi jantung, serta kelas BPJS yang digunakan BPJS tingkat 1. Adapun kelompok yang mempunyai nilai paling tinggi data pasien kelas BPJS adalah klaster 3 sebanyak 270 data pasien rawat inap peserta BPJS pada Usia 20-39 tahun dengan jenis penyakit asam urat, serta kelas BPJS yang digunakan BPJS tingkat 1. Penulis menyampaikan saran diantaranya: supaya interface yang dihasilkan bisa lebih berkembang lagi dengan menampilkan perhitungan langsung pada program yang ditemukan, serta supaya bisa menambah kriteria atau variabel lainnya, sehingga bisa diketahui korelasi yang lebih jelas lagi antara variabel-variabel tersebut[16].

Penelitian Aliman Darsani dkk di tahun 2021, melakukan pengelompokan data-data pengunjung/pasien di Puskesmas Ngkeran Kutacane sesuai dengan karakteristik yang diinginkan, dengan menggunakan algoritma *Clustering K-Means* yang diintegrasikan dengan aplikasi pemrograman berbasis dekstop. Peneliti menyimpulkan bahwa didapatkan sistem yang dirancang dan menghasilkan kelompok data pasien menjadi 3 klaster. Klaster 1 terdapat 20 data pasien, klaster 2 terdapat 13 data pasien serta klaster 3 terdapat 45 data pasien[17].

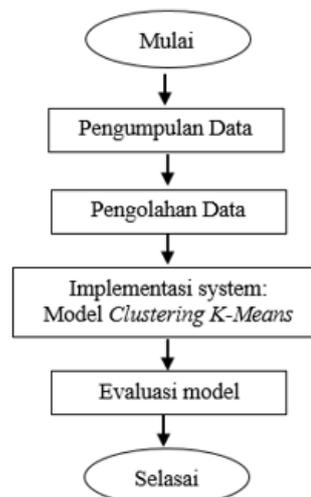
Penelitian Hendra Nusa Putra dan Dinda Putri Anisa pada tahun 2021 yang berjudul Klasterisasi Data Rekam Medis pada Diagnosa Penyakit Berdasarkan Usia Pasien Menggunakan Algoritma K-Means di Puskesmas Lubuk Alung. Penelitian ini bertujuan untuk mengetahui klasterisasi data rekam medis dengan memakai data kuantitatif deskriptif. Populasi yang digunakan adalah data kunjungan pasien 3 bulan terakhir pada tahun 2020 dari rekam medis pasien, pengumpulan data dilakukan dengan observasi serta analisis data dengan teknik/metode algoritma *clustering k-means*. Hasil yang diperoleh adalah peneliti menentukan atau mengelompokkan menjadi 3 klaster, yaitu klaster Penyakit Diderita Rendah sebanyak 318 data pasien (45%), Penyakit Diderita Sedang sebanyak 292 data pasien (41%), serta klaster Penyakit Diderita Tinggi sebanyak 318 data pasien (45%). Adapun penyakit terbanyak yang ditemukan di Puskesmas tersebut adalah Diabetes, Skizoprenia, Hipertensi, Stroke, Kelainan Refraksi, Rematik, Tinea, Asam Urat, Epilepsi dan Katarak. Metode klasterisasi bisa digunakan sebagai metode yang diharapkan bisa mempermudah penghitungan data penyakit pasien di Puskesmas Lubuk Alung[8].

Berdasarkan uraian tersebut diatas, maka akan dilakukan penelitian pengelompokan pengunjung/ pasien rawat jalan berdasarkan variabel umur, status kepesertaan BPJS kesehatan dan alamat pasien menggunakan algoritma clustering k-means di Puskesmas Aikmel Utara Kabupaten Lombok Timur. Hasil penelitian ini diharapkan dapat dapat memberikan kontribusi pengetahuan dalam pengelompokan data pasien dan pembentukan pola penyakit pasien rawat jalan di Puskesmas, dan dapat membantu mengelompokkan pasien rawat jalan berdasarkan karakteristik dan pola-pola kunjungan/ penyakit atau karakteristik kesehatan tertentu, sebagai bahan pengambil keputusan untuk menyediakan layanan yang lebih tepat dan spesifik sesuai dengan kebutuhan setiap kelompok pasien, serta sebagai acuan untuk pengembangan penelitian selanjutnya dalam penerapan algoritma data mining dalam bidang pelayanan kesehatan.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Jenis penelitian yang digunakan pada penelitian ini adalah penelitian kuantitatif deskriptif, yaitu penelitian yang melakukan analisis data drngan mendeskripsikan atau mengilustrasikan data yang sudah dikumpulkan sebelumnya sesuai ketentuan yang berlaku[18]. Adapun tahapan-tahapan yang dilaksanakan pada penelitian ini dapat dilihat pada gambar 1 berikut:



Gambar 1. Tahapan Alur Penelitian

Berdasarkan gambar 1 tersebut yang merupakan tahapan-tahapan dalam penelitian yang dilakukan oleh penulis sebagai acuan kegiatan adalah 4 tahap, yaitu tahap pengumpulan data, pengolahan data, tahap implementasi sistem, dan tahap evaluasi model.

**2.2 Pengumpulan Data**

Pengumpulan data pada penelitian ini menggunakan data skunder, yakni data/atau informasi yang dikumpulkan oleh orang lain selain penulis aslinya[18], dengan menggunakan teknik wawancara dan observasi melalui register kunjungan pasien rawat jalan pada periode Januari s/d Maret 2024. Adapun jenis data yang dikumpulkan yaitu data pengunjung atau pasien rawat jalan denan atribut jenis kelamin, usia pasien, status kepesertaan BPJS kesehatan, alamat pasien, serta diagnosa penyakit pasien.

**2.3 Pengolahan Data**

Pengolahan data yaitu melakukan proses transformasi atau coding pada data yang sudah disiapkan dengan menggunakan metode yang sudah ditentukan, yaitu data jenis kelamin, usia pasien, status kepesertaan BPJS dan pekerjaan pasien, serta diagnosa penyakit pasien dengan menggunakan algoritma clustering. Melakukan proses dalam rangka menyiapkan data yang siap dianalisis, meliputi:

- a. Dataset
 

Melakukan pemilihan dan pelabelan data berdasarkan atribut yaitu jenis kelamin, usia pasien, status kepesertaan BPJS, alamat pasien, serta diagnose pasien.
- b. Preprocessing data
 

Proses pengolahan data mentah, yaitu mempersiapkan data sesuai dengan kebutuhan analisis atau model yang akan digunakan. Pre processing data melibatkan pembersihan data, transformasi, dan penggabungan data.
- c. Validasi data
 

Proses menyiapkan dan memastikan data yang digunakan sudah siap dilakukan analisis menggunakan algoritma *clustering k-means*. Hasil proses validasi data sesuai pada tabel 1 berikut.

**Tabel 1.** Atribut Dataset Pasien Rawat Jalan

No.	Atribut	Deskripsi	Value	Inisialisasi
1.	Jenis kelamin	Jenis kelamin pasien	Laki-laki Perempuan	1 2
2.	Kelompok Umur	Kelompok Usia pasien (tahun),	Anak (0-9) Remaja (10-18) Dewasa (19-59) Lansia (=> 60)	1 2 3 4
3.	Status kepesertaan BPJS	Status peserta BPJS pasien	Peserta BPJS Tidak/Umum	1 2
4.	Alamat	Alamat desa peserta	Aikmel Utara Toya Aik Prapa Kb. Kerang Daya Luar wilayah	1 2 3 4 5
5.	Diagnosa	Jenis penyakit pasien berdasarkan ICD X	A00 A000 A001 ... Z99	1 2 3 ... dst

Berdasarkan tabel 1 tersebut, diperoleh masing-masing sudah dilakukan atau diberikan inisialisasi sehingga data siap dilakukan analisis, yaitu atribut jenis kelamin menjadi 2, kelompok umur menjadi 4, status kepesertaan BPJS menjadi 2, alamat menjadi 5, serta diagnosa penyakit sesuai dengan jumlah jenis penyakit sesuai diagnosa.

**2.4 Implementasi Data mining**

Melakukan proses pengambilan pola dari data yang tersedia sesuai dengan langkah-langkah algoritma *K-Means Clustering*, serta juga dengan bantuan tools *RapidMiner*. *Rapidminer* sendiri merupakan software yang berdiri sendiri yang ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi[19].

Langkah-langkah melakukan clustering dengan metode k-means adalah :

- a. Menentukan nilai/jumlah klaster (k)
- b. Menentukan centroid awal, titik-titik sampel yang menjadi anggota klaster.

$$C_i = \min + \frac{(i-1) * (max-min)}{n} + \frac{(max-min)}{2 * n} \tag{1}$$

- c. Menentukan nilai centroid atau titik terdekat dari centroid awal
- d. Menghitung square error untuk tiap klaster, yang merupakan jumlah kuadrat dari jarak euclidean antara tiap sampel dalam klaster dan titik tengahnya (centroid).

- e. Mengelompokkan kembali semua sampel berdasarkan jarak minimum dari masing-masing pusat, sehingga diperoleh distribusi baru dari sampel sesuai klasternya.
- f. Menjumlahkan keseluruhan error dari k-klaster
- g. Membandingkan hasil perhitungan dengan square error
- h. Mengulangi dari langkah ke-3 (jika hasilnya > error), sampai ditemukannya nilai square error turun signifikan.
- i. Melakukan redistribusi sampel ke dalam kelas yang sesuai dengan hasil perhitungan jarak terpendek tiap sampel[20].

**2.5 Evaluasi model**

Proses mengubah pola menjadi pengetahuan dengan tools *RapidMiner* sehingga dapat dimanfaatkan untuk mendukung pengambilan keputusan berupa pengelompokan data atau *cluster*. Proses ini menyajikan hasil dari scenario penelitian. Evaluasi dilakukan untuk mengetahui kualitas validitas cluater dengan menggunakan *Davies-Bouldin Indeks* (DBI). Hasil penelitian ini berupa fakta yang dihasilkan dari proses analisis algoritma *clustering k-means*.

Adapun langkah-langkah evaluasi pemodelan clustering k-means[20]:

- a. Menghitung *sum of square within cluster* (SSW): untuk mengetahui matriks kohesi dalam sebuah cluster ke-*i* :

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \tag{2}$$

- b. Menghitung *sum of square between cluster* (SSB): untuk mengetahui separasi antar cluster:

$$SSB_{ij} = d(c_i, c_j) \tag{3}$$

- c. Menghitung rasio (*R<sub>ij</sub>*) untuk mengatahui nilai perbandingan antara cluster ke-*i* dan cluater ke-*j* :

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \tag{4}$$

- d. Menghitung nilai *davies-bouldin index* (DBI)

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{i,j} \tag{5}$$

**3. HASIL DAN PEMBAHASAN**

**3.1 Dataset Penelitian**

Data yang digunakan pada penelitian ini ialah data yang didapatkan dari register kunjungan rawat jalan Puskesmas Aikmel Utara bulan Januari s/d Maret 2024. Jumlah dataset yang didapatkan sebanyak 1889 data pasien, dengan 5 atribut yaitu jenis kelamin, kelompok usia, status kepesertaan BPJS Kesehatan, alamat tempat tinggal, dan diagnose penyakit. Dataset hasil pengumpulan dan seleksi data dapat dilihat pada tabel 2 berikut.

**Tabel 2.** Dataset Hasil Pengumpulan Data

Pasien.	Jenis kelamin	Kelomok Usia	itatus Kepesertaan BPJS	Alamat tempat tinggal	Diagnosa penyakit
P1	Laki-laki	Dewasan (19-59)	Peserta BPJS	Toya	K008
P2	Laki-laki	Lansia (=> 60)	Peserta BPJS	Toya	K05
P3	Perempuan	Anak (0-9)	Tidak Peserta	Luar Wilayah	J188
P4	Perempuan	Dewasa (19-59)	Peserta BPJS	Aikmel Utara	Z30
P5	Perempuan	Dewasa (19-59)	Peserta BPJS	Kb. Kerang Daya	E119
...	...	...	...	...	...
P1888	Perempuan	Dewasa (19-59)	Peserta BPJS	Toya	J06
P1889	Perempuan	Lansis (=> 60)	Peserta BPJS	Aik Prapa	J06

Berdasarkan tabel 2 tersebut, dataset hasil pengumpulan data, diperoleh sebanyak 1889 data pasien dengan atribut jenis kelamin, usia, status kepeserta BPJS, alamat, serta diagnosa penyakit pasien.

**3.2 Pengolahan Data**

Pada tahap ini adalah tahap proses melakukan pengolahan data mentah, yaitu mempersiapkan data sesuai dengan kebutuhan analisis atau model yang akan digunakan. Pre processing data melibatkan pembersihan data, transformasi, dan penggabungan data, serta validasi data data yang digunakan sudah siap dilakukan analisis menggunakan algoritma clustering k-means. Pada proses pembersihan data ridak ditemukan *missing value*, *noise data*, tidak ada data yang tidak memiliki nilai atau keterangan. Dataset hasil *preprocessing* dan validasi data sebagaimana pada tabel 3 berikut.

**Tabel 3.** Dataset hasil processing dan validasi

Pasien.	Jenis kelamin (X1)	Klp umur (X2)	Status BPJS (X3)	Alamat (X4)	Diagnosa (X5)
P1	1	3	1	2	1

P2	1	4	1	2	2
P3	2	1	2	5	3
P4	2	3	1	1	4
P5	2	3	1	4	5
....	...	...	...	...	...
P1233	2	3	1	2	21
P1234	2	4	1	3	21

Berdasarkan tabel 3 tersebut, bahwa dari 1889 data pasien yang diperoleh, semuanya lolos dari proses *preprocessing* (pembersihan data, transformasi, dan penggabungan data) dan validasi data.

### 3.3 Implementasi/pengujian Clustering K-Means

- a. Menentukan jumlah kluster (k = 2)
- b. Jumlah centroid awal, masing-masing atribut:

Jenis kelamin:

$$C_1 = 1 + (1 - 1) * \frac{(2-1)}{2} + \frac{(2-1)}{2*2} = 1,25$$

$$C_2 = 1 + (2 - 1) * \frac{(2-1)}{2} + \frac{(2-1)}{2*2} = 1,75$$

Kelompok umur:

$$C_1 = 1 + (1 - 1) * \frac{(4-1)}{2} + \frac{(4-1)}{2*2} = 1,75$$

$$C_2 = 1 + (2 - 1) * \frac{(4-1)}{2} + \frac{(4-1)}{2*2} = 3,25$$

Status kepesertaan BPJS:

$$C_1 = 1 + (1 - 1) * \frac{(2-1)}{2} + \frac{(2-1)}{2*2} = 1,25$$

$$C_2 = 1 + (2 - 1) * \frac{(2-1)}{2} + \frac{(2-1)}{2*2} = 1,75$$

Alamat pasien:

$$C_1 = 1 + (1 - 1) * \frac{(5-1)}{2} + \frac{(5-1)}{2*2} = 2,00$$

$$C_2 = 1 + (2 - 1) * \frac{(5-1)}{2} + \frac{(5-1)}{2*2} = 4,00$$

Diagnosa penyakit:

$$C_1 = 1 + (1 - 1) * \frac{(237-1)}{2} + \frac{(237-1)}{2*2} = 60$$

$$C_2 = 1 + (2 - 1) * \frac{(237-1)}{2} + \frac{(237-1)}{2*2} = 178$$

- c. Menentukan jarak ke centroid awal:

Untuk P1:

$$C_1 = \sqrt{(1,25 - 1)^2 + (1,75 - 3)^2 + (2 - 2)^2 + (1,25 - 1)^2 + (60 - 1)^2} = 59,014$$

$$C_2 = \sqrt{(1,75 - 1)^2 + (3,25 - 3)^2 + (4 - 2)^2 + (1,75 - 1)^2 + (178 - 1)^2} = 177,015$$

Untuk P2:

$$C_1 = \sqrt{(1,25 - 1)^2 + (1,75 - 4)^2 + (2 - 1)^2 + (1,25 - 2)^2 + (60 - 2)^2} = 58,045$$

$$C_2 = \sqrt{(1,75 - 1)^2 + (3,25 - 4)^2 + (4 - 1)^2 + (1,75 - 2)^2 + (178 - 2)^2} = 176,016$$

Untuk P3:

$$C_1 = \sqrt{(1,25 - 2)^2 + (1,75 - 1)^2 + (2 - 2)^2 + (1,25 - 5)^2 + (60 - 3)^2} = 57,094$$

$$C_2 = \sqrt{(1,75 - 2)^2 + (3,25 - 1)^2 + (4 - 2)^2 + (1,75 - 5)^2 + (178 - 3)^2} = 175,018$$

dan seterusnya sampai P1889

- d. Menghitung square error untuk tiap kluster:

Hasil untuk tiap kluster dan masing-masing atribut hasilnya > 0

- e. Mengelompokkan data berdasarkan jarak terdekat: diperoleh (hasil perhitungan awal: C1=1676 dan C2=213

- f. Menjumlahkan total error dari kluster dan membandingkan dengan square error

- g. Karena hasil total error masih tinggi, maka diulangi dengan menggunakan centroid baru yaitu dari hasil perhitungan sampai total error menjadi nol atau hasil kluster sama dengan pusat kluster.

Pada penelitian ini, total error menjadi nol literasi dilakukan sampai 9 kali literasi, dan diperoleh (hasil perhitungan akhir: C1=1570 dan C2= 319. Hasil pengelompokan dengan 2 kluster seperti pada tabel 4 berikut.

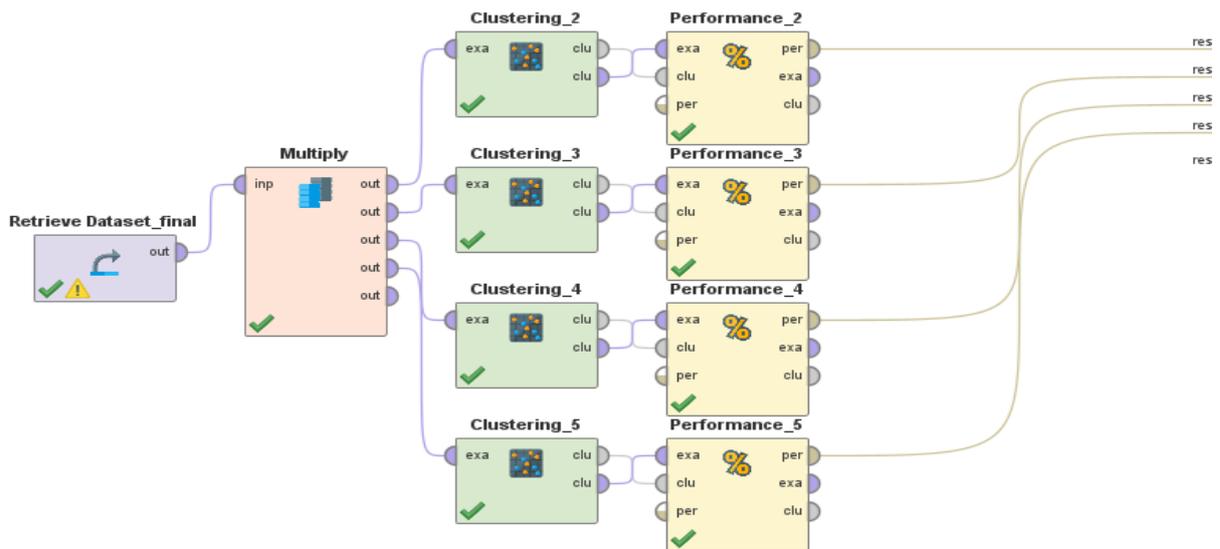
**Tabel 4.** Hasil pengelompokan kluster (k=2)

Pasien	Jenis kelamin (X1)	Klp umur (X2)	Status BPJS (X3)	Alamat (X4)	Diagnosa (X5)	C1	C2	Hasil
P1	1	3	1	2	1	27,69	147,86	1
P2	1	4	1	2	2	26,73	146,87	1
P3	2	1	2	5	3	25,85	145,89	1
P4	2	3	1	1	4	24,72	144,87	1

P5	2	3	1	4	5	23,74	143,87	1
...	...	...	...	...	...	...	...	...
P1888	2	3	1	2	21	7,717	127,86	1
P1889	2	4	1	3	21	7,845	127,87	1

Berdasarkan pada tabel 4 tersebut, semua data pasien (sebanyak 1889) sudah dilakukan pengelompok berdasarkan jenis kelamin, kelompok umur, status kepesertaan BPJS, alamat pasien, dan diagnosa penyakit pasien, menjadi 2 klaster

Perhitungan ini juga dilakukan dengan 3 klaster 4, dan 5 klaster. Klasterisasi ini juga dilakukan dengan menggunakan aplikasi “RapidMiner” dengan 2 klaster, 3 klaster, 4 klaster dan 5 klaster. Gambar 2 berikut merupakan gambar proses pengklasteran dengan tools RapidMiner.



Gambar 2. Operator Clustering

### 3.4 Evaluasi Model

a. Menghitung *sum of square within cluster* (SSW):

$$SSW1 = \frac{1}{1570} \sum_1^{1570} d(Xj, C1) = 15,768$$

$$SSW2 = \frac{1}{319} \sum_1^{319} d(Xj, C1) = 37,017$$

b. Menghitung *sum of square between cluster* (SSB):

$$SSB_{1,2} = d(c_1, c_2) = SSB_{2,1} = 120,191$$

c. Menghitung rasio (Rij)

$$R_{1,2} = R_{1,2} = \frac{15,768+37,017}{120,191} = 0,439$$

d. Menghitung nilai davies-bouldin index (DBI):

$$DBI = (0,439+0,439) / 2 = 0,439$$

Hasil evaluasi model (DBI) untuk masing-masing pengklasteran baik secara manual maupun dengan aplikasi RapidMiner adalah pada tabel 5 berikut:

Tabel 5. Hasil evaluasi DBI

Klaster (C)	Hasil DBI	Keterangan
2	0,439	
3	0,511	
4	0,477	
5	0,453	

Dari tabel 5 tersebut, bahwa setelah dilakukan evaluasi model dengan DBI terhadap hasil masing-masing pengklasteran diperoleh nilai DBI yang paling mendekati nol adalah yang dengan 2 klaster.

### 3.5 Interpretasi model

Berdasarkan hasil evaluasi DBI tersebut di atas, maka didapatkan nilai k yang optimal yaitu klaster dengan k = 2, karena memiliki nilai k yang paling mendekati nol yaitu dengan nilai DBI 0,439, sehingga dapat disimpulkan bahwa pada penelitian ini menunjukkan bahwa penggunaan 2 klaster merupakan klaster terbaik. Hasil masing-masing clustering atribut pada masing-masing klaster seperti pada tabel 6 berikut.

**Tabel 6.** Hasil klusterisasi pasien rawat jalan Puskesmas

Klaster dan jumlah	Data atribut
C1 jumlah 1570 data	Jumlah pasien perempuan 1074 dan pasien laki-laki 496 Jumlah pasien peserta BPJS Kesehatan 819 dan pasien bukan peserta BPJS 751 Kelompok umur didominasi oleh usia Dewasa sebanyak 883 (56,24%) Pengunjung rawat jalan terbanyak dari desa Toya sebanyak 488 (31,08%) Diagnosa penyakit pasien rawat jalan terbanyak J06 (ISPA akut) sebanyak 223 (14,20%)
C1 jumlah 319 data	Jumlah pasien perempuan 205 dan laki-laki 114 Jumlah pasien peserta BPJS Kesehatan 202 dan pasien bukan peserta BPJS 117 Kelompok umur didominasi oleh usia Dewasa sebanyak 191 (59,87%) Pengunjung rawat jalan terbanyak dari desa Toya sebanyak 108 (33,86%) Diagnosa penyakit pasien rawat jalan terbanyak Z36 (Pemeriksaan Kehamilan) sebanyak 29 (9,09%)

## 4. KESIMPULAN

Berdasarkan hasil pengolahan dan analisis penerapan algoritma *clustering k-means* pasien rawat jalan di Puskesmas Aikmel Utara Kabupaten Lombok Timur, maka diperoleh kesimpulan yaitu pengelompokan pasien rawat jalan dengan atribut jenis kelamin, kelompok usia, status kepesertaan BPJS, alamat pasien, dan diagnosa penyakit pasien menggunakan *clustering k-means* dengan jumlah 1889 data pasien, menunjukkan bahwa jumlah klaster yang dibentuk sebanyak 2 klaster, dengan jumlah data masing-masing klaster yaitu klaster 1 sebanyak 1570 data dan klaster 2 sebanyak 319 data. Klaster 1 didominasi oleh pasien perempuan sebanyak 1074 (68,4%), pasien peserta BPJS sebanyak 819 (52,2%), dengan kelompok usia terbanyak usia dewasa sebanyak 883 (56,2%), dan terbanyak dari desa Toya sebanyak 488 (31,1%), serta diagnose penyakit terbanyak ISPA akut (J06) sebanyak 223 (14,2%). Sedangkan klaster 2 didominasi oleh pasien perempuan sebanyak 205 (64,3%), pasien peserta BPJS sebanyak 202 (63,3%), dengan kelompok usia terbanyak usia dewasa sebanyak 191 (59,9%), dan terbanyak dari desa Toya sebanyak 108 (33,9%), serta diagnose penyakit terbanyak pemeriksaan kehamilan (Z34) sebanyak 29 (9,1%). Dari hasil klaster ini diperoleh informasi bahwa pengunjung rawat jalan di Puskesmas Aikmel Utara Kabupaten Lombok Timur terbanyak adalah dari desa Toya dan didominasi kelompok usia dewasa dan peserta BPJS kesehatan, serta penyakit terbanyak adalah ISPA akut, sehingga diharapkan dapat membantu pihak Puskesmas dalam pengambilan keputusan atau kebijakan yang berkaitan dengan program-program kesehatan di wilayah kerjanya. Peneliti juga menyarankan kepada peneliti berikutnya yang akan mengambil penelitian serupa supaya menambah atribut serta menggunakan algoritma lainnya.

## UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak Puskesmas Aikmel Utara yang telah mendukung terlaksananya penelitian ini, telah memberikan data kunjungan pasien rawat jalan Puskesmas.

## REFERENCES

- [1] Kemkes. RI, *Peraturan Menteri Kesehatan Nomor 31 tahun 2019 tentang Sistem Informasi Puskesmas*. Jakarta, 2019.
- [2] Kemkes. RI, *Peraturan Menteri Kesehatan Nomor 24 tahun 2022 tentang Rekam Medis*. Jakarta, 2022.
- [3] WHO, "International Classification Of Diseases and Related Health Problems (Klasifikasi Statistik Internasional Mengenai Penyakit Dan Masalah Kesehatan Terkait)," Volume 2 I.2005.
- [4] Dinas Kesehatan. Lombok Timur, *Profil Kesehatan Kabupaten Lombok Timur Tahun 2023*. 2024.
- [5] Puskesmas. aikmel Utara, *Laporan Tahunan Puskesmas Aikmel Utara Kabupaten Lombok Timur*. Aikmel, 2024.
- [6] Kusriani and E. T. Luthfi, *Algoritma Data Mining*. Surabaya: CV. ANDI OFFSET (Penerbit Andi), 2009.
- [7] M. A. dan M. Nasir, *Data Mining. Algoritma dan Implementasi*. Yogyakarta: CV. ANDI OFFSET (Penerbit Andi), 2020.
- [8] H.N. Putra and D.P. Anisa, "Klusterisasi Data Rekam Medis Pada Diagnosa Penyakit Berdasarkan Usia Pasien Menggunakan Algoritma K-Means di Puskesmas Lubuk Alung," vol. 3, no. 5, pp. 128–133, 2021.
- [9] T.A. Susanto, "Kelebihan dan Kekurangan Jenis-Jenis Clustering dalam Data Science," 2023. <https://blog.algorit.ma/jenis-clustering/#apa-itu-clustering>
- [10] N.K. Zuhail, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," vol. 1, pp. 200–205, 2022.
- [11] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk

- Pengelompokan Kasus Covid-19 di Dunia,” *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 18, no. 2, pp. 206–211, 2021.
- [12] M. Zubair, M. A. Iqbal, A. Shil, M. J. M. Chowdhury, M. A. Moni, and I. H. Sarker, “An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling,” *Ann. Data Sci.*, 2022, doi: 10.1007/s40745-022-00428-2.
- [13] J. W. & Sons, *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Education Services, 2015.
- [14] D. Ariyanto and I. Resercher, “Jurnal Sistim Informasi dan Teknologi Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut,” vol. 4, pp. 1–3, 2022, doi: 10.37034/jsisfotek.v4i1.117.
- [15] L. Masyfufah and I. Artikel, “Clustering Guna Menunjang Program Jaminan Balong Bendo Sidoarjo *Clustering BPJS's Inpatients By Using K -Means Clustering Method to Support The National Health Insurance Program At Anwar Medika Hospital Balong Bendo Sidoarjo*,” pp. 8–22, 2021.
- [16] N. Karolina, “Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering ( Studi Kasus : RSUD . Bangkatan ),” pp. 47–53, 2021.
- [17] A. Darsani, M. Yetri, and M.G. Suryanata, “Penerapan Data Mining Untuk Mengelompokkan Pasien Pada Puskesmas Ngkeran Kutacane Menggunakan Algoritma K-Means Clustering,” vol. 4, no. 9, pp. 1–13, 2021.
- [18] H. Ahyar *et al.*, *Buku Metode Penelitian Kualitatif & Kuantitatif*, no. March. 2020.
- [19] D.I.M. Reza Gustrianda, “Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan,” vol. 6, pp. 27–34, 2022, doi: 10.30865/mib.v6i1.3294.
- [20] I. Wahyudin, E. T. Tosida, and F. Andria, *Teori dan Panduan Praktis Data Science dan Big Data*, Pertama. Bogor: Lembaga Penelitian dan Pengabdian pada Masyarakat, Universitas Pakuan, 2019.