

Implementation of Clustering Method Using K-Means Algorithm for Grouping BPJS Health Patient Medical Record Data

Anggri Sapitri^{1*}, Nurdin^{2**}, Yesy Afrilia^{3*}

* Teknik Informatika, Universitas Malikussaleh, Aceh, Indonesia

anggri.200170146@mhs.unimal.ac.id¹, nurdin@unimal.ac.id², yesy.afrillia@unimal.ac.id³

Article Info

Article history:

Received 2025-07-02

Revised 2025-08-06

Accepted 2025-08-10

Keyword:

*K-Means Clustering,
Z-Score,
Medical Records,
BPJS Health,
RSUD Simeulue.*

ABSTRACT

Clustering medical record data of BPJS Health patients is essential in supporting data-driven decision-making in hospitals. This study aims to implement the K-Means algorithm to cluster patient medical records at RSUD Simeulue based on BPJS class and patient address variables. The data were first normalized using the Z-Score method to standardize variable scales, followed by the iterative application of the K-Means algorithm until convergence was reached at the sixth iteration. The study employed three Cluster, namely Cluster 1 (Very Many), Cluster 2 (Many), and Cluster 3 (Not Many). The final results show that Cluster 1 contains 258 patients from Class 1 and 292 from Class 2; Cluster 2 consists of 296 patients from Class 2; and Cluster 3 includes 101 patients from Class 1, 115 from Class 2, and 148 from Class 3. In addition to classification by BPJS class, clustering based on patient address revealed a dominant distribution from Simeulue Timur, Teluk Dalam, and Teupah Selatan sub-districts. The clustering results were implemented into a web-based information system using the Laravel framework and MySQL database, enabling hospital administrators to visualize and analyze patient data effectively. This study demonstrates that the K-Means algorithm can be effectively applied in classifying medical record data to support healthcare management decision-making.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Di masa digital saat ini, data rekam medis rumah sakit sangat besar dan kompleks dengan mencakup data tentang penderita diagnosis, pengobatan, serta riwayat medis yang ada, akan tetapi pengelolaan data rekam medis yang masih menjadi tantangan untuk rumah sakit. Pengelolaan data rekam medis salah satu bagian penting dari pengelolaan rumah sakit karena data ini merupakan informasi penting dalam memberikan layanan kesehatan kepada pasien [1]. Data mining adalah proses mengekstraksi pengetahuan bermanfaat dari kumpulan data berukuran besar dengan tujuan menemukan pola tersembunyi untuk mendukung pengambilan keputusan. Metode ini memanfaatkan teknik statistik, matematika, dan kecerdasan buatan, serta banyak diterapkan di bidang bisnis, sains, kesehatan, dan keuangan [2]. Di Kabupaten Simeulue sendiri proses pencatatan rekam medis masih menggunakan cara yang konvensional, yaitu dengan menulis pada kertas sehingga data yang ada pada rekam medis hanya bisa digunakan untuk melihat riwayat

kesehatan pasien. Dengan adanya teknologi bisa membantu proses pencatatan rekam medis menjadi lebih mudah dan efisien, yaitu dengan menggunakan computer [3].

Rekam medis hal wajib bagi dokter rumah sakit yang melakukan tindakan medis terhadap pasien sesuai dengan ketentuan, sehingga tidak ada alasan bagi dokter untuk tidak membuat rekam medis. Begitu pula dengan layanan medis. Pelayanan medis menjadi salah satu hal yang perlu ditingkatkan [4]. Dengan bertambahnya jumlah pelayanan kesehatan maka pembelajaran tentang rekam medis menjadi sangat penting, walaupun masih banyak masyarakat yang masih belum memahami tentang rekam medis. Rekam medis adalah suatu berkas yang berisi tentang rekam medis suatu rumah sakit dan didalamnya memuat informasi yang cukup banyak, karena setiap pasien pergi ke rumah sakit atau memeriksakan kesehatannya maka menjadi rekam medis yang menampung data-data yang menyebabkan berbagai masalah umum, seperti ruang penyimpanan yang besar, hilangnya data pasien yang diperlukan, hilangnya informasi dan lain-lain [5].

Penelitian terdahulu oleh Amalia fahada pemanfaatan web sig untuk pemetaan dan klasterisasi jenis hasil perikanan tangkapan menggunakan metode K-Means klustering, jenis ikan pelagis 31 ikan tergolong ke dalam kategori C1, 4 ikan tergolong ke dalam kategori C2 dan 1 ikan tergolong ke dalam kategori C3. Untuk jenis ikan damersial 13 ikan tergolong ke dalam kategori C1, 17 ikan tergolong ke dalam kategori C2 dan 2 ikan tergolong ke dalam kategori C3. Untuk jenis ikan karang 2 ikan tergolong ke dalam kategori C1, 3 ikan tergolong ke dalam kategori C2 dan 1 ikan tergolong ke dalam kategori C3[6].

Penelitian oleh Annur penerapan data mining dalam upaya menentukan strategi penjualan variasi mobil, cluster 1 merupakan kelompok barang dengan harga yang rendah dan diminati konsumen, cluster 2 barang dengan harga sedang dan diminati konsumen, dan cluster 3 barang dengan harga tinggi dan tidak terlalu diminati konsumen [7].

Clustering adalah salah satu teknik dalam data mining yang bertujuan untuk mengelompokkan data yang serupa ke dalam kelompok-kelompok atau cluster berdasarkan kesamaan karakteristiknya. Dalam proses clustering, data diorganisir ke dalam beberapa kelompok sedemikian rupa sehingga objek dalam kelompok yang sama memiliki kesamaan yang tinggi, sedangkan objek antar kelompok memiliki perbedaan yang signifikan [8]. Z-Score adalah nilai statistik untuk menstandarkan data ke rata-rata nol dan deviasi standar satu, sehingga data pada skala berbeda bisa dibandingkan seimbang. Metode ini tidak mengubah distribusi, hanya menyesuaikan skala, dan dapat diterapkan dengan fungsi *StandardScaler()* dari scikit-learn sebelum analisis lanjutan seperti clustering [9].

Penelitian sebelumnya yang menggunakan algoritma K-Means clustering diantaranya yaitu dalam klasterisasi pasien BPJS di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo, Klaster 1 terdiri dari 91 perempuan (51%), klaster 2 terdiri dari 26 pasien perempuan (14%), dan klaster 3 terdiri dari 63 pasien laki-laki (35%). Pasien yang paling banyak sakit terdapat pada Kecamatan Krian dan balon bendo, dengan pasien jenis kelamin perempuan yang mendominasi, sedangkan untuk diagnosa penyakit berdasarkan ICD lebih banyak terdiagnosa born in hospital (Z38.0), Diarrgoea (A09+E86), dan Hemorrhagic (A91.5) [10].

Banyak metode pengelompokkan lain seperti pada penelitian Fuzzy C-Means merupakan Teknik pengumpulan data Dimana keberadaan setiap bagian data dalam suatu cluster di tentukan oleh derajat keanggotaannya. Fuzzy C-Means menerapkan pengelompokkan fuzzy dimana setiap data menjadi anggota beberapa cluster dengan derajat keanggotaan yang berbeda-beda di setiap cluster. Fuzzy C-Means merupakan algoritma iteratif yang menerapkan iterasi pada proses pengelompokkan data dan tujuan dari fuzzy sendiri untuk mendapatkan pusat dari cluster yang nantinya digunakan untuk mengetahui data yang masuk ke dalam cluster [11].

Kemudian pada penelitian lainnya yakni metode Naive Bayes Classifier (NBC) merupakan sebuah metode

pengelompokkan berdasarkan pada teorema bayes. algoritma Naive Bayes Classifier secara otomatis dapat melakukan proses pengklasifikasian data karya ilmiah atau tugas akhir, dan proses klasifikasi menjadi lebih akurat dengan menggunakan data latih dalam jumlah besar. Metode klasifikasi ini menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksikan peluang di masa depan berdasarkan pada pengalaman di masa lalu, sehingga metode ini lebih dikenal dengan Teorema Bayes Algoritma Naive Bayes Classifier efektif dalam mengklasifikasikan karya ilmiah, terbukti dari hasil pengujian data dari 20 karya ilmiah berdasarkan parameter latar belakang [12].

Penelitian ini mengumpulkan data COVID-19 dari 23 kabupaten/kota di Aceh dengan menggunakan 6 variabel yang terdiri dari terkonfirmasi, dalam perawatan, sembuh, meninggal, suspek, dan probable. Hasil penelitian klasterisasi penyebaran COVID-19 adalah sebagai berikut: Satu kabupaten/kota klaster 1 (zona merah), empat kabupaten/kota klaster 2 (zona kuning), delapan belas kabupaten/kota klaster 3 (zona hijau). Berdasarkan hasil penelitian ini, Algoritma Fuzzy C-Means dapat digunakan dan diaplikasikan dengan baik dalam klasterisasi penyebaran COVID- 19 di Provinsi Aceh [13].

Dengan adanya fasilitas jaminan sosial BPJS Kesehatan yang diselenggarakan pemerintah dengan menjunjung prinsip gotong royong dalam penyediaan jaminan kesehatan masyarakat, banyak pasien yang menggunakan fasilitas tersebut. Salah satunya RSUD Simeulue mengalami kendala dalam upaya peningkatan jenis fasilitas yang akan diprioritaskan bagi pasien yang menggunakan jaminan sosial BPJS Kesehatan. Pemilihan data rekam medis pasien untuk dianalisis karena terdapat informasi-informasi penting yang dapat memberikan pengetahuan baru bagi pihak manajemen rumah sakit ketika akan mengambil keputusan peningkatan fasilitas kesehatan Serta dapat diketahui pola penyebaran penyakit berdasarkan informasi pada data rekam medis dapat dilakukan dengan menggunakan metode Clustering berdasarkan atribut Nama, Alamat, Umur, Jenis Penyakit, Jenis Kelamin, Dan Kelas BPJS [14].

K-Means adalah suatu Teknik analisis kelompok yang bertujuan untuk membagi N objek pengamatan menjadi K kelompok berdasarkan mean (rata-rata) terdekat. Metode ini mirip dengan algoritma Expectation-Maximization untuk Gaussian Mixture juga mencari pusat kelompok dalam data melalui iterasi perbaikan [15]. Algoritma yang dapat digunakan untuk melakukan clustering data adalah metode K-Means clustering. Algoritma K-Means clustering merupakan sebuah teknik dalam Data Mining yang dapat mempartisi data menjadi beberapa cluster sehingga data dengan karakteristik yang sama akan dikelompokkan menjadi satu cluster sekaligus data yang mempunyai karakteristik berbeda akan dikelompokkan ke dalam cluster lain. K-Means melakukan dua proses, yaitu proses penentuan pusat cluster (centroid) dan proses pencarian anggota dari masing-masing cluster dan

memerlukan k parameter masukan atau banyaknya cluster [16].

BPJS Ketenagakerjaan adalah badan hukum publik yang memberikan perlindungan sosial bagi pekerja formal maupun nonformal di Indonesia. Programnya mencakup jaminan kecelakaan kerja, hari tua, pensiun, dan kematian [17]. Program jaminan kesehatan menyediakan layanan mulai dari pemeriksaan, pengobatan, rawat inap, operasi, hingga tindakan medis lainnya. Peserta membayar iuran bulanan sesuai kelas, dan dengan itu berhak mendapatkan pelayanan kesehatan yang ditanggung BPJS. Tujuan utama BPJS Kesehatan adalah untuk meningkatkan akses masyarakat Indonesia terhadap pelayanan kesehatan yang berkualitas tanpa harus mengalami beban finansial yang berat. Dengan demikian, BPJS Kesehatan diharapkan dapat memberikan perlindungan finansial bagi peserta dalam menghadapi risiko kesehatan serta meningkatkan kesejahteraan dan kualitas hidup masyarakat secara keseluruhan [18].

Dalam penelitian ini penulis menggunakan metode algoritma k -means untuk mengelompokkan data rekam medis dalam menghitung jumlah pasien dan data penyakit yang di alami pasien pengguna BPJS Kesehatan K -Means menawarkan kemudahan penggunaan, efisiensi, dan kemampuan untuk bekerja pada data berukuran besar dan penulis tertarik untuk melakukan penelitian untuk mengelompokkan dan mengelola data rekam medis [19]. Rekam medis atau ICD mencatat riwayat kesehatan pasien, termasuk diagnosis, hasil pemeriksaan, pengobatan, tindakan medis, serta informasi penunjang lain yang mendukung proses perawatan dan keputusan klinis [20].

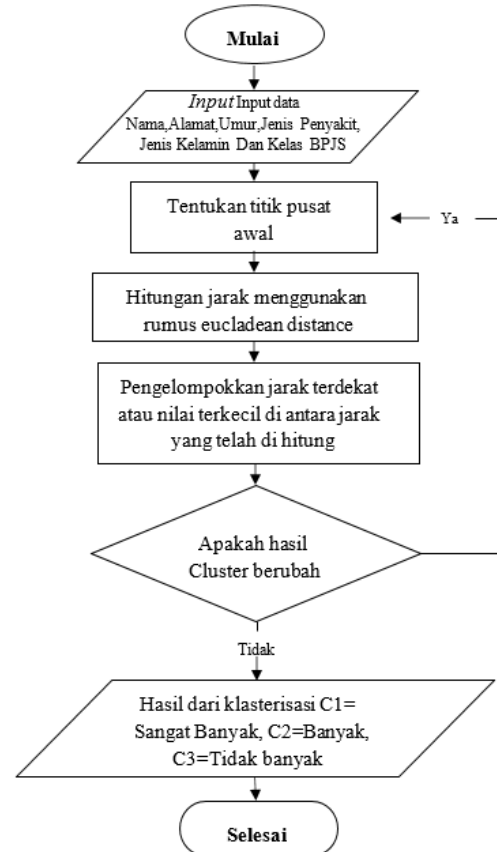
RSUD SIMEULUE merupakan salah satu rumah sakit yang terletak di Kabupaten Simeulue Rumah Sakit ini memiliki jumlah pasien yang banyak datang setiap bulannya, sehingga data riwayat pasien terakumulasi dalam rekam medis dan kegunaan data yang tersedia tidak terlacak. Meskipun demikian, data ini mempunyai potensi besar untuk memberikan wawasan berharga bagi para profesional kesehatan, peneliti, dan pembuat kebijakan berbasis bukti. Salah satu pendekatan untuk mengatasi tantangan ini adalah pengembangan data menggunakan teknik clustering K -Means. Selain itu, teknik pengelompokan K -Means menggunakan data rekam medis dapat digunakan untuk mendeteksi pola prevalensi penyakit di RSUD Simeulue.

Pentingnya penelitian ini adalah belum adanya model sistem klasterisasi data rekam medis pasien BPJS kesehatan di Kabupaten Simeulue. Dengan melakukan klasifikasi data rekam medis pasien BPJS kesehatan menggunakan algoritma K -means dengan pendekatan data mining sehingga dapat menentukan hasil pengelompokkan data rekam medis pasien BPJS kesehatan.

II. METODE

Metode penelitian ini digunakan untuk menjelaskan tahapan-tahapan yang dilakukan dalam pengumpulan dan pengolahan data rekam medis pasien BPJS Kesehatan di

RSUD Simeulue. Penelitian ini menerapkan algoritma K -Means sebagai metode clustering untuk mengelompokkan data berdasarkan atribut-atribut seperti nama, alamat, umur, jenis penyakit, jenis kelamin, dan kelas BPJS. Adapun tahapan pelaksanaan metode ini ditunjukkan pada Gambar 1.



Gambar 1. Skema K -Means

Gambar diatas merupakan Gambaran proses atau skema perancangan sistem pada penelitian ini, dari tahap di mulai hingga selesai:

1. Proses skema algoritma K -Means di mulai.
2. Proses penginputan data rekam medis yakni Nama, Alamat, Umur, Jenis Penyakit, Jenis Kelamin Dan Kelas BPJS.
3. Menentukan titik pusat awal yang diinginkan dengan cara mengurutkan data dari yang terkecil hingga terbesar pada setiap variabel yang di ambil dari data 1 tahun terakhir yakni tahun 2023 kemudian dibagi menjadi 3 klompok pada setiap cluster ...kemudian memilih masing-masing titik pusat kelompok yakni Cluster 1 (Pasien Perkotaan dengan Tingkat Pemanfaatan Tinggi), Cluster 2 (Pasien Pinggiran dengan Tingkat Pemanfaatan Sedang), dan Cluster 3 (Pasien Pedesaan dengan Tingkat Pemanfaatan Rendah). Jumlah cluster ($k = 3$) ditentukan menggunakan Metode Elbow dengan cara memplot nilai *Within-Cluster Sum of Squares* (WCSS) terhadap beberapa nilai k , kemudian dipilih

titik siku pada grafik sebagai jumlah cluster yang optimal. Penentuan ini juga dikonfirmasi dengan menghitung Silhouette Score untuk memastikan kualitas pemisahan antar cluster dan kekompakan di dalam cluster.

- Setelah menentukan titik pusat awal maka akan dilakukan perhitungan jarak *euclidean* dengan rumus pada persamaan (1)

$$Z=(X-\mu)/\sigma \dots\dots\dots(1)$$

Dimana :

- Z = Z-Score
- X = Nilai asli
- μ = Rata-rata
- σ = Standar deviasi kolom

- Pengelompokkan dilakukan berdasarkan jarak terdekat atau jarak terkecil dari hasil perhitungan jarak *euclidean* pada tiap data.

- Hasil pengelompokkan menggunakan *K-Means* berhenti apabila hasil klastering berubah. apabila hasil klastering berubah maka menghitung titik pusat baru dengan rumus pada persamaan (2).

Adapun rumus untuk menghitung jarak *Euclidean* adalah sebagai berikut:

$$D(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \dots\dots\dots(2)$$

Keterangan:

D= titik dokumen

x_i = Data kriteria

μ_j = *centroid* pada cluster ke- j

Kemudian menghitung jarak *euclidean* Kembali dan mengelompokkan berdasarkan jarak terkecil.

- Jika perhitungan sudah selesai maka akan keluar hasil klaster yang dapat menentukan data tersebut masuk pada klaster sangat banyak, banyak atau Tidak banyak.

- Proses perhitungan selesai.

Proses pengolahan data dilakukan menggunakan bahasa pemrograman Python untuk perhitungan algoritma *K-Means*. Selain itu, sistem informasi dikembangkan menggunakan *Laravel* sebagai *framework web* dan *MySQL* sebagai basis data.

III. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan tiga cluster utama dari data pasien BPJS di RSUD Simeulue menggunakan *K-Means*. Data dinormalisasi dengan metode *Z-Score* agar skala variabel seimbang, dan proses iterasi dilakukan 6 kali hingga cluster stabil.

A. Analisis Data

Penelitian ini menganalisis hubungan variabel data dengan pengelompokan pasien BPJS Kesehatan di RSUD Simeulue menggunakan metode *K-Means Clustering*. Data sebanyak 1.210 rekam medis pasien (350 Kelas 1, 500 Kelas 2, 360 Kelas 3) diolah melalui normalisasi untuk meningkatkan akurasi. Hasil clustering digunakan untuk melihat proporsi

tiap kelas BPJS dan pola yang terbentuk. Berikut ini adalah data variabel dan data skala atau data parameter untuk menentukan data yang akan dinormalisasi:

TABEL I
DATA KELAS BPJS

Kelas	Nilai
BPJS Kelas 1	1
BPJS Kelas 2	2
BPJS Kelas 3	3

TABEL II
DATA BPJS DI RSUD SIMEULUE

No	Nama	Alamat	Umur	Jenis Kelamin	Kelas	Id penyakit
1	Pasien 1	Simeulue Timur	4	Pria	1	1
2	Pasien 2	Simeulue timur	70	Wanita	1	1
3	Pasien 3	Simeulue Timur	2	Pria	3	1
4	Pasien 4	Simeulue Timur	28	Pria	3	2
5	Pasien 5	Simeulue Timur	13	Pria	2	3
6	Pasien 6	Simeulue Timur	31	Pria	1	4
7	Pasien 7	Simeulue Timur	49	Pria	3	4
8	Pasien 8	Simeulue Timur	65	Wanita	2	5
9	Pasien 9	Simeulue Timur	71	Wanita	1	6
10	Pasien 10	Simeulue Timur	8	Pria	2	1
...
...
...
...
...
1206	Pasien 1206	Teupah Selatan	33	Pria	3	28
1207	Pasien 1207	Simeulue Timur	45	Pria	3	264
1208	Pasien 1208	Teluk Dalam	45	Pria	3	10
1209	Pasien 1209	Teluk dalam	25	Wanita	3	265
1210	Pasien 1210	Teluk Dalam	48	Wanita	2	266

B. Normalisasi Data Menggunakan *Z-Score*

Dalam proses pengolahan data menggunakan algoritma *K-Means*, tahap normalisasi merupakan langkah penting untuk menyetarakan skala dari masing-masing atribut yang digunakan, terutama karena data yang digunakan dalam penelitian ini memiliki karakteristik skala yang berbeda-beda,

seperti alamat pasien yang direpresentasikan dalam bentuk numerik dan kelas BPJS yang dikategorikan ke dalam tiga kelas. Oleh karena itu, normalisasi dilakukan menggunakan metode **Z-Score** agar data yang diolah memiliki nilai rata-rata 0 dan standar deviasi 1, sehingga tidak ada atribut yang mendominasi proses perhitungan jarak. Proses normalisasi ini bertujuan untuk meningkatkan akurasi dalam pengelompokan serta mengurangi bias akibat skala data yang tidak seimbang.

Untuk melakukan normalisasi Z-Score menggunakan persamaan 1.

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Dimana :

- Z = Z-Score
- X = Nilai asli
- μ = Rata-rata
- σ = Standar deviasi kolom

$$Z \text{ Alamat Pasien 1} = \frac{X - \mu}{\sigma} = \frac{1 - 3,226}{2,904} = -0,76$$

$$Z \text{ Kelas Pasien 1} = \frac{X - \mu}{\sigma} = \frac{1 - 2,07}{0,812} = -1,31$$

$$Z \text{ Alamat Pasien 1191} = \frac{X - \mu}{\sigma} = \frac{10 - 3,226}{2,904} = 2,33$$

$$Z \text{ Kelas Pasien 1191} = \frac{X - \mu}{\sigma} = \frac{1 - 2,07}{0,812} = -1,31$$

Perhitungan dilakukan pada seluruh data BPJS RSUD Simeulue sehingga di dapatkan nilai Z-Score dari setiap data seperti pada tabel berikut :

TABEL III
NILAI Z-SCORE BPJS RSUD SIMEULUE

No	Nama	Z-Score	
		ZX	ZY
1	pasien1	-0.76	-1.31
2	pasien 2	-0.76	-1.31
3	pasien 3	-0.76	1.14
4	pasien 4	-0.76	1.14
5	pasien 5	-0.76	-0.08
6	pasien 6	-0.76	-1.31
7	pasien 7	-0.76	1.14
8	pasien 8	-0.76	-0.08
9	pasien 9	-0.76	-1.31
10	pasien 10	-0.76	-0.08
11	pasien 11	-0.76	-0.08
12	pasien 12	-0.76	1.14
13	pasien 13	-0.76	-1.31
14	pasien 14	-0.76	1.14
15	pasien 15	-0.76	1.14
16	pasien 19	-0.76	-0.08

17	pasien 23	-0.76	-0.08
18	pasien 24	-0.76	1.14
19	pasien 26	-0.76	-1.31
...
1209	pasien 1191	2.33	-1.31

Keterangan :

ZX = Z-Score Alamat

ZY = Z-Score Kelas

C. Hasil Iterasi K-Means

Proses K-Means dilakukan dalam 6 iterasi. Setiap iterasi menghasilkan perubahan posisi centroid, hingga iterasi ke-6 di mana hasil klasterisasi stabil dan tidak mengalami perubahan lagi. Berikut adalah ringkasan hasil iterasi:

- **Iterasi 1-5:** Terjadi perpindahan anggota antar cluster akibat pergeseran centroid.
- **Iterasi 6:** Tidak terdapat perubahan anggota cluster, yang menandakan proses telah konvergen.

Tabel 1 menunjukkan hasil akhir centroid dari setiap cluster, sedangkan Tabel 2 menampilkan jumlah pasien per kelas BPJS di masing-masing cluster.

TABEL IV
CENTROID BARU UNTUK ITERASI 6

Centroid Baru Berdasarkan pemetaan cluster hasil iterasi sebelumnya		
Centroid	ZX (alamat)	ZY (kelas)
1	1.3922253	0.0747527
2	-0.6049662	1.1400000
3	-0.5914000	-0.6569818

Dan Dilakukan Kembali perhitungan jarak untuk iterasi ke 6 menggunakan centroid baru yang telah terbentuk.

Iterasi 6

$$d(1,1) = \sqrt{((-0.76) - 1.39)^2 + ((-1.31) - 0.07)^2} = 2,55$$

$$d(1,2) = \sqrt{((-0.76) - (-0.60))^2 + ((-1.31) - 1.14)^2} = 2,45$$

$$d(1,3) = \sqrt{((-0.76) - (-0.59))^2 + ((-1.31) - (-0.65))^2} = 0,67$$

$$d(1191,1) = \sqrt{((2,33) - 1.39)^2 + ((-1.31) - 0.07)^2} = 1,67$$

$$d(1191,2) = \sqrt{((2,33) - (-0.60))^2 + ((-1.31) - 1.14)^2} = 3,82$$

$$d(1191,3) = \sqrt{((2,33) - (-0.59))^2 + ((-1.31) - (-0.65))^2} = 2,99$$

Dan di dapatkan hasil di iterasi ke 6 seperti pada tabel berikut :

TABEL V
HASIL ITERASI 6

No	Nama	C1	C2	C3	Cluster
1	pasien 1	2.559	2.455	0.674	3
2	pasien 2	2.559	2.455	0.674	3
3	pasien 3	2.401	0.155	1.805	2
4	pasien 4	2.401	0.155	1.805	2
5	pasien 5	2.158	1.23	0.601	3
6	pasien 6	2.559	2.455	0.674	3
7	pasien 7	2.401	0.155	1.805	2
8	pasien 8	2.158	1.23	0.601	3
9	pasien 9	2.559	2.455	0.674	3
10	pasien 10	2.158	1.23	0.601	3
11	pasien 11	2.158	1.23	0.601	3
12	pasien 12	2.401	0.155	1.805	2
13	pasien 13	2.559	2.455	0.674	3
14	pasien 14	2.401	0.155	1.805	2
15	pasien 15	2.401	0.155	1.805	2
16	pasien 19	2.158	1.23	0.601	3
17	pasien 23	2.158	1.23	0.601	3
18	pasien 24	2.401	0.155	1.805	2
19	pasien 26	2.559	2.455	0.674	3
...
1209	pasien 1191	1.672	3.823	2.993	1

Hasil perhitungan jarak di iterasi 6 telah konsisten dengan iterasi sebelumnya maka perhitungan jarak dapat di hentikan.

D. Analisis Hasil Clustering

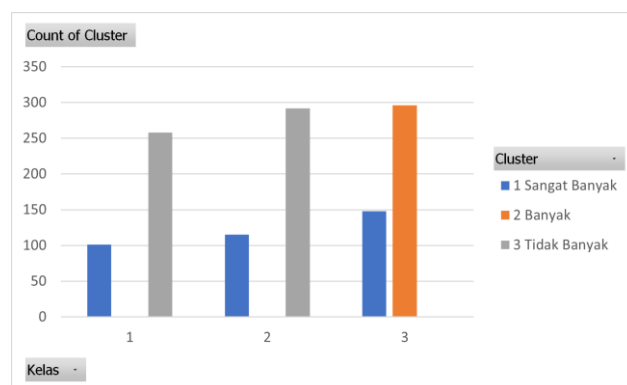
Berdasarkan hasil perhitungan jarak pada iterasi ke-6, data pasien berhasil dikelompokkan ke dalam tiga cluster. Distribusi pasien berdasarkan kelas BPJS menunjukkan bahwa pada Cluster 1 terdapat 148 pasien dengan kelas BPJS 3, 115 pasien dengan kelas BPJS 2, dan 101 pasien dengan kelas BPJS 1. Pada Cluster 2, seluruh pasien berasal dari kelas BPJS 2 dengan jumlah sebanyak 296 pasien. Sementara itu, Cluster 3 terdiri atas 292 pasien dengan kelas BPJS 2 dan 258 pasien dengan kelas BPJS 1.

Selanjutnya, jika dikaji berdasarkan alamat atau wilayah asal pasien, pada Cluster 1 terdapat 18 pasien dari Kecamatan Simeulue Barat, 114 pasien dari Kecamatan Teluk Dalam, 108 pasien dari Kecamatan Teupah Selatan, 55 pasien dari Kecamatan Simeulue Tengah, 8 pasien dari Kecamatan Alafan, dan 61 pasien dari Kecamatan Salang. Adapun pada Cluster 2, mayoritas pasien berasal dari Kecamatan Simeulue Timur sebanyak 235 pasien, disusul oleh Kecamatan Teupah Tengah sebanyak 19 pasien, Kecamatan Simeulue Cut sebanyak 22 pasien, Kecamatan Teupah Barat sebanyak 9 pasien, dan Kecamatan Simeulue Barat sebanyak 11 pasien. Sementara itu, Cluster 3 mencakup 412 pasien dari

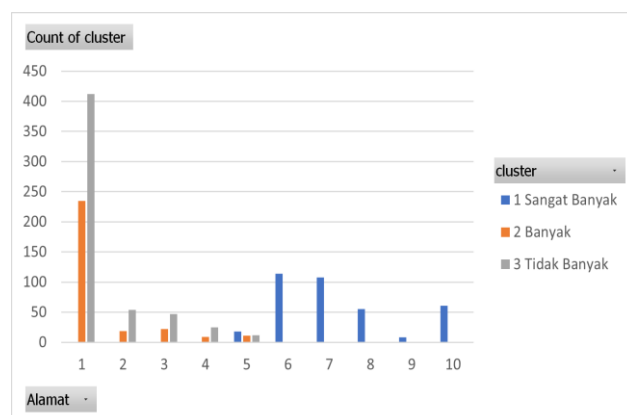
Kecamatan Simeulue Timur, 54 pasien dari Kecamatan Teupah Tengah, 47 pasien dari Kecamatan Simeulue Cut, 25 pasien dari Kecamatan Teupah Barat, dan 12 pasien dari Kecamatan Simeulue Barat.

Temuan ini menunjukkan adanya pola persebaran pasien yang cukup bervariasi berdasarkan kelas BPJS dan wilayah asalnya, yang dapat menjadi dasar untuk analisis lebih lanjut terkait karakteristik kelompok pasien. Setiap cluster mewakili kelompok pasien dengan karakteristik berbeda. Cluster 1 didominasi oleh pasien BPJS Kelas 2 dari pusat kota dengan permintaan layanan yang tinggi. Cluster 2 berisi pasien dari wilayah pinggiran dengan tingkat pemanfaatan layanan yang sedang. Cluster 3 mencakup pasien dari daerah pedesaan dengan frekuensi kunjungan rumah sakit yang lebih rendah namun kelas BPJS yang lebih beragam. Pola ini menunjukkan adanya hubungan antara domisili pasien dan kelas BPJS.

Untuk memvalidasi kualitas hasil clustering, dilakukan perhitungan nilai Silhouette Coefficient dengan rata-rata skor sebesar 0,61 yang menunjukkan pemisahan cluster yang baik. Selain itu, Davies-Bouldin Index juga digunakan untuk validasi internal dengan nilai 0,54 yang menunjukkan kekompakan cluster yang memadai. Visualisasi PCA (Principal Component Analysis) juga dilakukan untuk mengonfirmasi pemisahan cluster pada ruang dua dimensi. Untuk lebih jelasnya dapat dilihat pada gambar grafik di bawah ini :



Gambar 1. Grafik clustering berdasarkan Kelas BPJS



Gambar 2. Grafik berdasarkan Alamat

E. Implementasi Sistem

Hasil dari proses pengelompokan data pasien BPJS Kesehatan menggunakan algoritma K-Means telah diimplementasikan ke dalam sebuah sistem informasi berbasis web. Sistem ini dikembangkan menggunakan framework Laravel yang fleksibel dan modern, serta menggunakan MySQL Sistem web yang dikembangkan memiliki fitur dashboard yang menampilkan ringkasan hasil cluster, peta sebaran pasien (heatmap), dan grafik untuk analisis cluster. Pengguna juga dapat mengekspor laporan hasil pengelompokan pasien dalam format PDF dan Excel. Pengujian usability sederhana dilakukan dengan melibatkan 5 staf administrasi rumah sakit untuk mengevaluasi antarmuka sistem, proses input data, dan fitur pembuatan laporan. Masukan dari pengguna digunakan untuk menyempurnakan tampilan dashboard dan fungsi ekspor laporan.

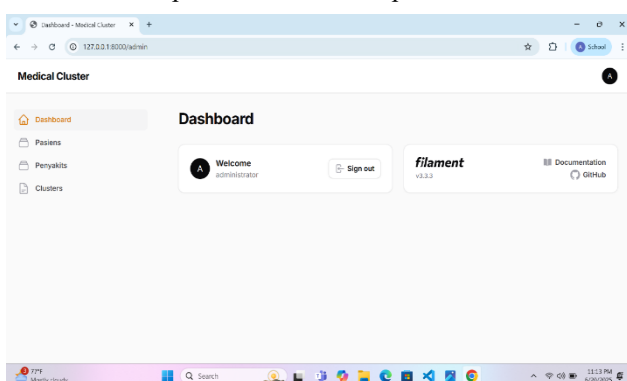
Selain itu, sistem dilengkapi dengan fungsi filter untuk menampilkan data cluster berdasarkan kelas BPJS atau kecamatan domisili pasien, sebagai basis data untuk menyimpan dan mengelola data pasien secara terstruktur.

Sistem ini memiliki beberapa fitur utama, yaitu:

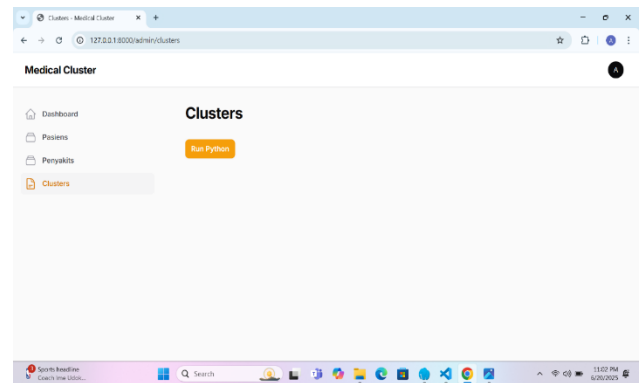
- Form Input Data Rekam Medis, yang memungkinkan petugas rumah sakit untuk memasukkan data pasien dengan atribut seperti nama, alamat, umur, jenis kelamin, jenis penyakit, dan kelas BPJS.
- Halaman Cluster, yang menampilkan hasil pengelompokan data dalam bentuk tabel klasifikasi dan grafik visualisasi (pie chart atau bar chart), berdasarkan hasil dari proses K-Means.
- Dashboard Sistem, yang menyajikan ringkasan data pasien dan informasi penting lainnya dalam format yang mudah dipahami dan responsif.

Dengan adanya visualisasi hasil pengelompokan, sistem ini membantu pihak manajemen rumah sakit dalam memahami distribusi pasien berdasarkan wilayah tempat tinggal dan tingkat layanan BPJS yang digunakan. Sistem ini juga dapat digunakan sebagai alat bantu dalam merumuskan kebijakan peningkatan layanan kesehatan, terutama dalam menentukan prioritas pelayanan dan perencanaan anggaran.

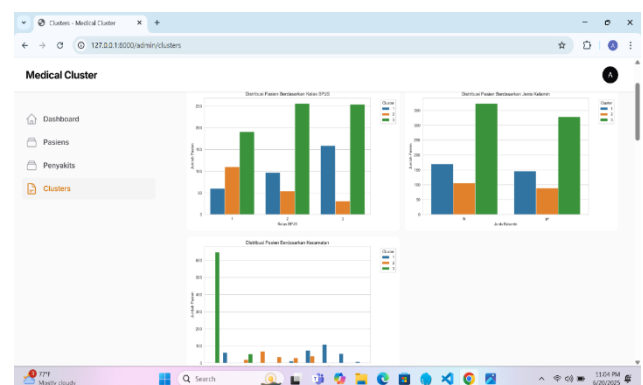
Berikut tampilan sistem hasil implementasi:



Gambar 3. Tampilan Dashboard Sistem



Gambar 4. Tampilan Cluster Sistem



Gambar 5. Tampilan Hasil Pengelompokan Pasien

Sistem ini diharapkan menjadi **solusi digital yang** komprehensif dalam mendukung pengelolaan data rekam medis di rumah sakit, khususnya dalam hal analisis dan visualisasi data pasien BPJS Kesehatan. Dengan adanya sistem informasi berbasis web ini, proses pengambilan keputusan strategis oleh pihak manajemen rumah sakit dapat dilakukan dengan lebih cepat, tepat, dan berbasis pada data yang valid. Selain itu, sistem ini juga memberikan kemudahan dalam mengakses informasi secara real-time, mempercepat proses pelaporan, dan mengurangi risiko kesalahan manusia (human error) dalam pencatatan serta klasifikasi data pasien. Tampilan antarmuka yang interaktif dan user-friendly juga memungkinkan pengguna non-teknis, seperti staf administrasi dan tenaga medis, untuk berinteraksi dengan sistem secara mudah tanpa memerlukan pelatihan teknis mendalam.

Dengan kemampuan klasifikasi otomatis menggunakan algoritma K-Means, sistem ini tidak hanya sekadar menyimpan data, tetapi juga berfungsi sebagai alat bantu analitik yang dapat digunakan untuk mengidentifikasi tren, pola penyakit, serta distribusi layanan kesehatan berdasarkan wilayah dan kelas sosial. Hal ini sangat penting dalam membantu rumah sakit menetapkan kebijakan pelayanan yang berbasis kebutuhan pasien yang sebenarnya. Secara keseluruhan, implementasi sistem ini menjadi bagian penting

dalam transformasi digital rumah sakit menuju sistem informasi kesehatan yang lebih modern, akuntabel, dan berbasis teknologi data mining. Sistem ini juga berpotensi untuk dikembangkan lebih lanjut di masa mendatang, seperti integrasi dengan data pasien non-BPJS, perluasan atribut klasifikasi, dan penggabungan dengan sistem rekam medis elektronik lainnya.

IV. KESIMPULAN

Berdasarkan hasil penelitian ini dapat di simpulkan bahwa Metode K-Means berhasil membagi data pasien BPJS menjadi tiga cluster yang optimal, didukung pendekatan Elbow Method. Hasil clustering menunjukkan pola hubungan yang jelas antara domisili pasien dan kelas BPJS. Model clustering ini dapat mendukung perencanaan sumber daya dan peningkatan layanan kesehatan di rumah sakit. Sistem yang dibangun mempermudah visualisasi data dan pembuatan laporan untuk mendukung pengambilan keputusan manajemen.

DAFTAR PUSTAKA

- [1] 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, hal. 158–168, 2023, doi: 10.37600/tekinkom.v6i1.857.
- [2] N. Cahyana dan A. Aribowo, "Metode Data Mining K-Means Untuk Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat," *Semin. Nas. Inform. Medis*, no. 5, hal. 24–31, 2018.
- [3] D. Andreswari, R. Efendi, dan K. Prastio, "Clustering Data Rekam Medis untuk Penentuan Penyakit Endemi di Daerah Kabupaten Bengkulu Selatan dengan Mengimplementasikan Metode FUZZY C-MEANS," *J. Rekursif*, vol. 11, no. 1, hal. 42–52, 2023, [Daring]. Tersedia pada: <http://ejournal.unib.ac.id/index.php/rekursif/42>
- [4] W. Aulia, A. Putera Utama Siahaan, L. Marlina, dan M. Iqbal, "Analisis Algoritma K-Means Clustering Dalam Identifikasi Tingkat Risiko Penyakit Berdasarkan Data Rekam Medis Pasien," *J. Sci. Soc. Res.*, vol. 4307, no. 3, hal. 3457–3465, 2025, [Daring]. Tersedia pada: <http://jurnal.goretanpena.com/index.php/JSSR>
- [5] Nanda, dkk., "Analisis Data Mining Untuk Klasterisasi Data Rekam Medis Menggunakan Algoritma K-Means Pada Rumah Sakit Sylvani Binjai," *Indones. J. Educ. Comput. Sci.*, vol. 1, no. 3, hal. 82–88, 2023.
- [6] M. Faisal, N. Nurdin, F. Fajriana, dan Z. Fitri, "Information and Communication Technology Competencies Clustering For Students For Vocational High School Students Using K-Means Clustering Algorithm," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 3, hal. 111–120, 2022, doi: 10.52088/ijesty.v2i3.318.
- [7] J. Wandana, S. Defit, dan S. Sumijan, "Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means," *J. Inf. dan Teknol.*, vol. 2, hal. 4–9, 2020, doi: 10.37034/jidt.v2i4.73.
- [8] D. Dahnil, "Implementation of K-Means Clustering Method to Lecturers Based on Publications of National Journals and Accredited Sinta," *JEECS (Journal Electr. Eng. Comput. Sci.)*, vol. 8, no. 1, hal. 27–40, 2023, doi: 10.54732/jeees.v8i1.4.
- [9] D. Prasetyawan, A. Mulyanto, dan R. Gatra, "Pemetaan Lintasan Karir Alumni Berdasarkan Analisis Cluster: Kombinasi K-Means dan Reduksi Dimensi Autoencoder," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, hal. 198–207, 2025, doi: 10.29408/edumatic.v9i1.29713.
- [10] Y. Saputra Sy, "Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means," *J. Sistim Inf. dan Teknol.*, vol. 5, hal. 33–37, 2022, doi: 10.37034/jsisfotek.v5i2.162.
- [11] N. Nurdin, S. Fitriani, Z. Yunizar, dan B. Bustami, "Clustering the Distribution of COVID-19 in Aceh Province Using the Fuzzy C-Means Algorithm," *JTAM (Jurnal Teor. dan Apl. Mat.)*, vol. 6, no. 3, hal. 665, 2022, doi: 10.31764/jtam.v6i3.8576.
- [12] N. Nurdin, M. Suhendri, Y. Afrilia, dan R. Rizal, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC)," *Sistemasi*, vol. 10, no. 2, hal. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.
- [13] N. S.Kom., M.Kom (SCOPUS ID=ID: 57201646662), U. M. Putri Nasution, H. A.-K. Aidilof, dan B. Bustami, "Implementation of Fuzzy C-Means to Determine Student Satisfaction Levels in Online Learning," *Sistemasi*, vol. 11, no. 1, hal. 121, 2022, doi: 10.32520/stmsi.v11i1.1638.
- [14] N. N. Dzikrulloh dan B. D. Setiawan, "Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri)," vol. 1, no. 5, hal. 378–385, 2017.
- [15] F. P. Ferdy Pangestu, N. Y. Nur Yasin, R. C. Ronald Chistover Hasugian, dan Y. Yunita, "Penerapan Algoritma K-Means Untuk Mengklasifikasi Data Obat," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, hal. 53–62, 2023, doi: 10.32736/sisfokom.v12i1.1461.
- [16] M. S. Sasmita, "Implementasi Algoritma K-Means Untuk Clustering Data Penyakit Di Puskesmas Kotagede 2 Yogyakarta," *SUBMIT J. Ilm. Teknol. Infomasi dan Sains*, vol. 4, no. 1, hal. 1–9, 2024, doi: 10.36815/submit.v4i1.3195.
- [17] Nurdiana dan Yusrizal, "Analisis Persepsi Pelayanan BPJS Ketengakerjaan Di Tanjung Morawa," *J. Soc. Sci. Res.*, vol. 3, no. 2, hal. 10344–10352, 2023.
- [18] H. Thabrany, "Bpjsnsionalataubpjsd Badan Penyelenggara Jaminan Kesehatan Nasional : Sebuah Policy Paper dalam Analisis Kesesuaian," hal. 15, 2015.
- [19] 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, hal. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.
- [20] R. Ordila, R. Wahyuni, Y. Irawan, dan 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, hal. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.