Vol. 7 No.01 2025
E-ISSN: 2685-6921
Universitas Ma'arif Nahdlatul Ulama Kebumen

Klasterisasi Pasien Rawat Inap BPJS pada RS Islam Assyifa Sukabumi menggunakan Metode *K-Means*

Irfan Nafis Sjamsuddin[1*], Dasya Arief Firmansyah[2], Yuni Laferani[1]

[1]*Universitas Siliwangi, Tasikamalaya, Indonesia*

[2]*Universitas Bina Sarana Informatika, Jakarta Pusat, Indonesia*

irfansjam@unsil.ac.id[*]

Received: 19/05/2025 | *Revised: 03/05/2025* | *Accepted: 07/06/2025*

Abstrak

Rumah sakit sebagai penyedia layanan kesehatan berkualitas menghadapi tantangan dalam pengelolaan data pasien yang semakin kompleks. Data tersebut belum dimanfaatkan secara optimal oleh manajemen rumah sakit serta memiliki potensi besar untuk dianalisis dan menjadi dasar dalam pengambilan keputusan. Bentuk optimaliasi tersebut dapat berupa pemanfaatan data pasien BPJS rawat inap di Rumah Sakit Islam Assyifa Sukabumi pada periode triwulan keempat 2024 menggunakan teknik *data mining*. Teknik yang diusulkan dalam penelitian ini adalah metode *K-Means* untuk mengelompokkan pasien BPJS berdasarkan variabel-variabel tertentu seperti usia, jenis kelamin, diagnosa penyakit, kelas rawat inap dan lama rawat inap. Adapun hasil penelitian ini mengungkapkan bahwa terdapat 3 klaster dari 3526 data pasien. Klaster 1 terdiri dari 1545 pasien dengan penyakit menular yang disebabkan oleh mikroorganisme. Klaster 2 terdiri dari 712 pasien dengan penyakit yang berkaitan dengan masalah kehamilan, kelahiran maupun gejala yang harus dii dentifikasi melalui pemeriksaan klinikal atau laboratorium lanjutan. Klaster 3 terdiri dari 1269 pasien dengan penyakit yang berkaitan dengan sistem pernafasan, sistem pencernaan, dan sistem sirkulasi darah. Evaluasi menunjukkan pengelompokkan pasien BPJS dengan hasil 3 klaster memiliki kualitas terbaik, dengan nilai Davies-Bouldin Index (DBI) sebesar 0,561. Hasil penelitian dapat menjadi acuan dalam merencanakan alokasi sumber daya rumah sakit. Saran untuk penelitian berikutnya adalah penerapan teknik data mining lainnya dalam optimalisasi pengelolaan data rumah sakit.

Kata kunci: BPJS, Klasterisasi, K-Means, Rumah sakit

Abstract

Hospitals as providers of quality health services face challenges in managing increasingly complex patient data. The data has not been optimally utilized by hospital management and has great potential to be analyzed and become a basis for decision-making. Optimization can be utilized by BPJS inpatient data at the Assyifa Sukabumi Islamic Hospital in the fourth quarter of 2024 using data mining

techniques. The technique proposed in this study is the K-Means method to group BPJS patients based on certain variables such as age, gender, disease diagnosis, inpatient class, and length of hospitalization. The results of this study revealed that there were 3 clusters of 3526 patient data. Cluster 1 consists of 1545 patients with infectious diseases caused by microorganisms. Cluster 2 consists of 712 patients with diseases related to pregnancy, childbirth, or symptoms that must be identified through further clinical or laboratory examinations. Cluster 3 consists of 1269 patients with diseases associated with the respiratory system, digestive system, and blood circulation system. The evaluation showed that the grouping of BPJS patients with 3 cluster results had the best quality, with a Davies-Bouldin Index (DBI) value of 0.561. The study results can be a reference in planning the allocation of hospital resources. Suggestions for further research are the application of other data mining techniques in optimizing hospital data management.

Keywords: BPJS, Clustering, K-Means, Hospital

1 Pendahuluan

Rumah sakit merupakan Fasilitas Kesehatan Rujukan Tingkat Lanjutan (FKRTL) yang menyelenggarakan pelayanan kesehatan secara lengkap, meliputi layanan rawat jalan, rawat inap, serta layanan kegawatdaruratan. Saat ini semua layanan tersebut telah terintegrasi dengan program jaminan kesehatan yang diselenggarakan oleh pemerintah yaitu BPJS (Badan Penyelenggara Jaminan Sosial) Kesehatan Indonesia. Dalam beberapa tahun terakhir (2014-2023), terjadi peningkatan yang signifikan (85,60%) keberadaan rumah sakit yang telah mengintegrasikan program BPJS dengan pelayanan kesehatan yang diberikan (BPJS Kesehatan, 2024). Peningkatan tersebut harus sejalan dengan upaya rumah sakit dalam meningkatkan efektifitas dalam memberikan pelayanan kesehatan. Salah satu aspek penting dalam proses pelayanan tersebut adalah penyelenggaraan dan pengelolaan rekam medis yang berbasis digital dan terintegrasi. Peraturan Menteri Kesehatan Indonesia Nomor 24 Tahun 2024 menyebutkan rekam medis eletronik merupakan dokumen yang dibuat dengan menggunakan sistem eletronik, yang mencakup data identitas, hasil pemeriksaan, tindakan pengobatan, dan berbagai layanan lain yang diberikan kepada pasien (Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 Tentang Rekam Medis, 2022).

Sesuai dengan Permenkes No 24 Tahun 2024, data kesehatan berupa rekam medis memiliki peran krusial dalam meningkatkan mutu pelayanan kesehatan di rumah sakit. Rumah Sakit Islam Assyifa Sukabumi sebagai salah satu FKTRL yang berfokus dalam menyediakan layanan jasa kesehatan yang berkualitas di kota Sukabumi dan memiliki visi untuk menjadi *Smart Hospital* (Firmansah et al., 2021) menghadapi tantangan dalam pengelolaan data pasien yang semakin kompleks, terutama data pasien layanan rawat jalan, rawat inap, maupun layanan kegawatdaruratan. Data tersebut belum dimanfaatkan secara optimal oleh manajemen rumah sakit serta memiliki potensi besar untuk dianalisis dan menjadi dasar dalam pengambilan keputusan. Data yang kompleks dapat dimanfaatkan secara maksimal, terutama dalam pengelompokkan atau segmentasi jenis penyakit maupun pasien berdasarkan karakteristik tertentu (Ningsih et al., 2025). Hasil pengelolaan data tersebut diharapkan dapat menjadi solusi dari berbagai masalah yang berpotensi muncul dalam manajemen pengelolaan rumah sakit seperti keterlambatan dalam mengidentifikasi tren penyakit maupun kurang optimalnya alokasi sumber daya (Herdiaman et

al., 2024). Salah satu pemanfaatan tersebut yang akan dicoba diterapkan yaitu pemanfaatan data pasien BPJS di RS Islam Assyifa Sukabumi menggunakan teknik *data mining*.

Klasterisasi atau *Clustering* adalah salah satu teknik data mining untuk mencari dan mengelompokkan data menjadi beberapa subset atau klaster berdasarkan kemiripan karakteristik (*similiarity*) antara satu data dengan data yang lain (Sulastri & Gufroni, 2017). Tujuan utama dari teknik ini yaitu mendistribusikan objek, orang, atau peristiwa ke dalam kelompok atau klaster di mana semakin mirip antar data maka akan berada dalam kluster yang sama dan semakin tidak mirip maka berada dalam klaster yang berbeda (Saputra, 2023). Dalam keilmuan data mining terdapat dua metode pengelompokkan dalam teknik *clustering*, yaitu *hierarchical clustering* dan *non-hierarchical*. Salah satu metode metode *non-hierarchical clustering* tersebut dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya) kemudian setelah diketahui jumlah klaster dilanjutkan proses klasterisasi, sering disebut sebagai *K-Means Clustering* (Sulastri & Gufroni, 2017). Metode tersebut akan digunakan sebagai metode pengelompokkan data dalam penelitian ini.

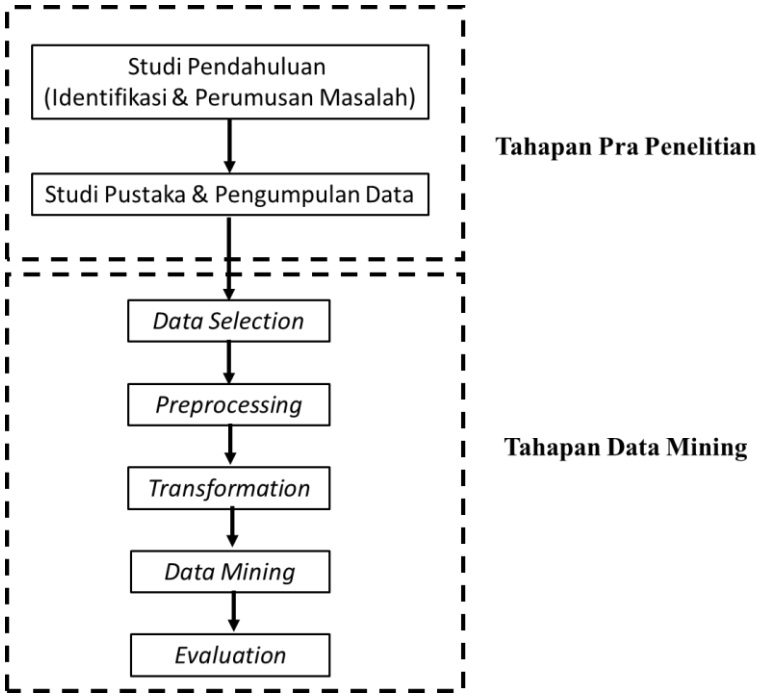
Beberapa penelitian terdahulu menunjukkan bahwa algoritma *K-Means* merupakan metode klasterisasi yang sederhana namun efektif dalam mengelompokkan data dalam jumlah besar berdasarkan kemiripan atribut. Penelitian (Jain, 2010) menyebutkan efektifitas metode ini dalam sektor kesehatan terutama dalam segmentasi pasien, pengelolaan sumber data, hingga perencanaan program kesehatan berbasis data. Salah satu penelitian terdahulu menerapkan *K-Means* dalam pengelompokkan pasien BPJS rawat inap pada periode 1 bulan dimana diperoleh hasil yaitu 3 klaster. Hasil tersebut juga dievaluasi memiliki kualitas terbaik dengan *Davies-Bouldin Index* (DBI) sebesar 0,406 (Herdiaman et al., 2024). Penelitian lainnya (Nengsih, 2025) berfokus pada segmentasi pasien berdasarkan pola kunjungannya, baik itu yang menggunakan layanan rawat jalan maupun rawat inap. Hasil penelitian tersebut juga diperoleh 3 klaster tetapi tidak dievaluasi hasil klasterisasi yang dihasilkannya.

Berbeda dengan penelitian sebelumnya, peneliti akan lebih berfokus pada pengelompokkan jenis penyakit berdasarkan data pasien BPJS rawat inap di RS Islam Assyifa Sukabumi pada periode bulan Oktober – Desember 2024 melalui penerapan algoritma *K-Means* berserta hasil evaluasinya. Hasil ini diharapkan dapat memberikan gambaran pola pasien yang lebih baik sehingga manajemen rumah sakit dapat melakukan intervensi yang lebih tepat sasaran. Hal tersebut juga dapat menunjang peningkatan efisiensi dan mutu layanan rumah sakit, serta khususnya pada bidang ilmu komputer dan manajemen kesehatan melalui penerapan metode data mining dalam pengelolaan rumah sakit.

2 Metodologi Penelitian

Penelitian ini mencakup serangkaian langkah atau tahapan yang dimulai dari awal sampai dengan akhir penelitian. Penelitian ini secara umum dapat dibagi menjadi dua tahapan utama. Tahapan tersebut diuraikan sebagai berikut (Gambar 1). Tahapan pra-penelitian, berupa studi pendahuluan yang meliputi identifikasi masalah dan perumusan masalah, kemudian dilanjutkan dengan studi pustaka terkait masalah tersebut, kemudian diakhiri dengan pengumpulan data pasien BPJS rawat inap di RS Islam Assyifa Sukabumi. Tahapan *data mining*, berupa proses untuk memahami apa yang yang harus dilakukan untuk mengolah data mentah menjadi pengetahuan yang berguna. Metode data mining yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Database* (KDD). Metode ini memiliki keunggulan karena tahapan

kegiatannya yang sederhana dan mudah dipahami. Metode ini memiliki 5 tahapan yaitu *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation* (Saputra, 2023).



Gambar 1 Tahapan Penelitian

2.1 Tahapan Pra Penelitian

a. Studi Pendahuluan (Identifikasi & Perumusan Masalah)

Tahap pertama yang dilakukan oleh peneliti ialah mengenali permasalahan yang berkaitan dengan isu yang akan diteliti melalui pencarian topik. Setelah menemukan masalah yang layak untuk dikaji, peneliti melanjutkan untuk menyusun rumusan masalah tentang Klasterisasi Pasien BPJS di RS Islam Assyifa Sukabumi.

b. Studi Pustaka & Pengumpulan Data

Pada tahap ini, dilakukan pendalaman terhadap literatur yang berkaitan dengan topik penelitian guna memperoleh pemahaman yang lebih komprehensif. Studi pustaka juga dimanfaatkan untuk mencari sumber pustaka yang relevan dan mendukung penelitian. Dalam penelitian ini, data yang akan digunakan adalah data rawat inap pasien BPJS di Rumah Sakit Islam Assyifa Sukabumi pada periode triwulan keempat (Oktober – Desember) tahun 2024. Data tersebut mencakup 3.540 entri dengan 14 atribut, yaitu nomor, nomor rekam medis, agama, jenis kelamin, diagnosa masuk, diagnosa primer, diagnosa ICD, diagnosa sekunder, kelas rawat inap, lama rawat inap, status pulang, usia, golongan usia, alamat, dan kota. Berikut ini adalah dataset yang akan dianalisis menggunakan *K-Means Clustering*.

2.2 Tahapan Data Mining

a. *Data Selection*

Memilih data berarti melihat kelayakan data tersebut untuk diproses dalam data mining (Saputra, 2023). Data yang layak atau baik adalah data yang jumlahnya banyak (kuantitas) dan berkualitas (data memiliki variasi yang beragam dan atribut yang unik).

b. *Preprocessing*

Preprocessing adalah tahap analisis data sebelum data tersebut siap digunakan, berfungsi untuk memastikan bahwa dataset yang akan digunakan tidak memiliki kekurangan ketika akan diproses menggunakan komputer. Tahapan ini secara umum meliputi *data cleaning*, *data transformation*, dan *data reduction*.

c. *Transformation*

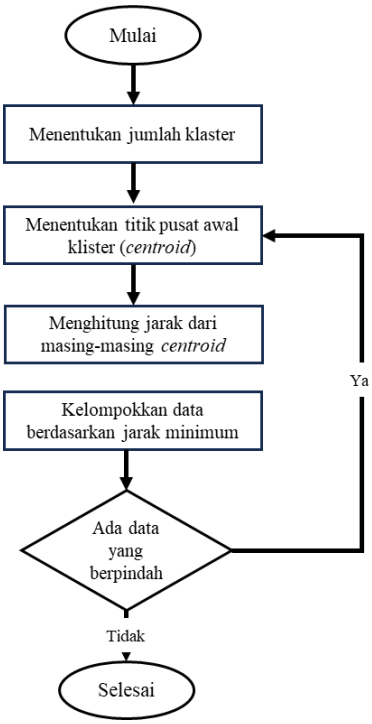
```
[10] s = (data.dtypes == 'object')
     object_cols = list(s[s].index)

     print("Categorical variables in the dataset:", object_cols)

     Categorical variables in the dataset: ['L/P', 'Diagnosa Primer', 'Kelas', 'Status Pulang', 'Golongan']

     LE_LP = LabelEncoder()
     LE_DP = LabelEncoder()
     LE_K = LabelEncoder()
     LE_SP = LabelEncoder()
     LE_G = LabelEncoder()

     data['L/P'] = LE_LP.fit_transform(data['L/P'])
     data['Diagnosa Primer'] = LE_DP.fit_transform(data['Diagnosa Primer'])
     data['Kelas'] = LE_K.fit_transform(data['Kelas'])
     data['Status Pulang'] = LE_SP.fit_transform(data['Status Pulang'])
     data['Golongan'] = LE_G.fit_transform(data['Golongan'])
```

Gambar 2 Contoh Penerapan Transformasi Data menggunakan Python

Transformation yaitu perubahan bentuk dimana pada tahapan ini, data akan dirubah sesuai dengan kebutuhan pada tahapan selanjutnya untuk menyesuaikan dengan tipe algoritma data mining yang akan digunakan. Dalam penerapannya, tahapan ini seringkali dilakukan sebelum atau bersamaan dengan *preprocessing* dengan alasan menyesuaikan dengan kebutuhan. Dalam penelitian ini, data akan dirubah menjadi data numerik menggunakan bahasa pemrograman *Python* karena menyesuaikan dengan algoritma *K-Means* (Gambar 2).

d. *Data Mining*

Secara sederhana, proses data mining dalam penelitian ini mengunakan algoritma *K-Means* (Saputra, 2023) dapat diuraikan dalam *flow chart* sebagai berikut (Gambar 3).



Gambar 3 Flow Chart Algoritma *K-Means*

1. Menentukan banyaknya cluster K
2. Menentukan titik pusat awal (centroid) setiap cluster secara acak
3. Menghitung jarak antara data dan centroid dengan menggunakan persamaan Euclian Distance (Sulastri & Acep, 2017) , dengan rumus sebagai berikut.

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \cdots + (x_{ki} - x_{kj})^2} \ldots (1)$$

Keterangan :
d (ij) = Jarak data ke i ke pusat cluster j
X ki = Data ke- i pada atribut data ke- k
X kj = Data ke- j pada atribut data ke- k

4. Mengelompokkan data ke dalam cluster dengan jarak yang paling dekat atau minimum dari setiap data dengan *centroid* tertentu dan memasukkan data ke dalam cluster tersebut
5. Menghitung nilai *centroid* baru dari setiap cluster yang telah terbentuk menggunakan rumus sebagai berikut.
6. Ulangi langkah ketiga sampai kelima sehingga sudah tidak ada lagi data yang berpindah ke cluster lain

e. *Evaluation*

Davies Bouldin Index merupakan metode evaluasi yang mengukur seberapa terpisah atau terpaut suatu cluster dari cluster lainnya, Nilai indeksnya berkisar antara 0 sampai ∞. Nilai yang semakin kecil menunjukkan model clustering tersebut lebih tepat, sedangkan nilai yang semakin besar menunjukkan bahwa model *clustering* tersebut kurang tepat (Saputra, 2023).

3 Hasil dan Pembahasan

Sebelum algoritma K-Means diterapkan dalam proses *data mining*, data terlebih dahulu akan melalui tahap seleksi yang merupakan langkah awal dalam proses *Knowledge Discovery in Database*. Pada tahap ini, data awal dikumpulkan sebanyak 3540 data dan kemudian dilakukan pemilihan variabel. Dalam tahap ini, dilakukan penghapusan variabel atau atribut yang tidak dibutuhkan dalam penelitian ini antara lain no, no RM, agama, diagnosa masuk, diagnosa diagnosa primer_ICD, diagnosa sekunder, alamat, kota, Langkah ini dilakukan untuk menyesuikan data dengan atribut atau variabel yang akan diproses.Setelah tahap seleksi data, langkah berikutnya adalah pembersihan data, yang melibatkan penghapusan duplikat, perbaikan data yang tidak konsisten, dan perbaikan kesalahan penulisan. Dalam tahap ini, dilakukan penghapusan data yang memiliki nilai kosong sebanyak 34 data. Diperoleh data terakhir sebanyak 3526 data (Tabel 1).

Tabel 1 Dataset setelah *Data Cleaning*.

No	L/P	Diagnosa Primer	Kelas	Lama Rawat Inap	Status Pulang	Usia	Golongan
1	P	O62.4	Kelas I	4	Pulang Hidup	27	Dewasa
2	P	O62.4	Kelas II	4	Pulang Hidup	28	Dewasa
3	L	J44.9	Kelas I	4	Pulang Hidup	59	Lansia
4	L	A49.9	Kelas II	4	Pulang Hidup	24	Remaja
5	P	O62.4	Kelas I	4	Pulang Hidup	24	Remaja
…	…	…	…	…	…	…	…
3526	L	N18.3	ICU	2	Meninggal kurang 48 jam	55	Lansia

Tahapan transformasi merupakan bagian krusial dalam mengolah data, di mana dalam tahapan ini beberapa data yang tidak berjenis numerik akan diransformasikan menjadi numerik menggunakan inisialisasi atau koding. Beberapa data yang dirubah antara lain jenis kelamin, diagnosa primer, kelas, status pulang, golongan. Gambaran tersebut dapat dilihat pada tabel berikut.

Tabel 2 Hasil Transformasi Data

No	L/P	DP	K	LRI	SP	U	G
1	1	231	3	4	4	27	3
2	1	231	4	4	4	28	3
3	0	118	3	4	4	59	4
4	0	7	4	4	4	24	6
5	1	231	3	4	4	24	6
…	…	…	…	…	…	…	…
3526	0	182	1	2	2	55	4

3.1 Perhitungan K-Means Clustering

Setelah semua data ditransformasikan sesuai dengan kebutuhan penelitian, tahapan selanjutnya adalah pengelompokkan data tersebut menggunakan algoritma *K-Means Clustering*. Langkah – langkah penerapan tersebut akan diuraikan sebagai berikut.

1. Langkah pertama yaitu menentukan jumlah klaster yang dibutuhkan dalam penelitian ini agar hasil yang diperoleh optimal. Berdasarkan beberapa hasil penelitian sebelumnya (Ali & Masyfufah, 2020; Herdiaman et al., 2024; Ningsih et al., 2025) pengelompokkan data pada penelitian ini akan dibagi menjadi 3 klaster.
2. Langkah kedua yaitu menentukan titik pusat awal untuk setiap klaster. Pada penelitian ini, titik pusat awal dipilih secara acak dan hasilnya berupa titik pusat awal atau *centroids* dari setiap klaster sebagai berikut (Tabel 3).

Tabel 3 Titik pusat awal (*Centroid* awal)

Centroid Awal	L/P	DP	K	LRI	SP	U	G
C1	0	7	4	4	4	24	6
C2	0	118	3	4	4	59	4
C3	1	29	4	3	4	22	6

3. Setelah memperoleh titik pusat awal (*centroids* awal) untuk setiap klaster, langkah selanjutnya yaitu melakukan perhitungan jarak antara setiap data dengan titik pusat awal setiap klaster menggunakan perhitungan jarak Euclidean. Berikut ini adalah hasil perhitungan tersebut dari setiap klaster.

   Perhitungan jarak Euclidean pada iterasi 1

   Berikut ini adalah beberapa hasil perhitungan jarak Euclidean pada klaster 1 :

   $$D(1,1) = \sqrt{(1-0)^2} + (231 - 7)^2 + (3 - 4)^2 + (4 - 4)^2 + (4 - 4)^2 + (27 - 24)^2 + (3 - 6)^2$$
   $$= 224{,}02$$
   $$D(2,1) = \sqrt{(1-0)^2} + (231 - 7)^2 + (4 - 4)^2 + (4 - 4)^2 + (4 - 4)^2 + (28 - 24)^2 + (3 - 6)^2$$
   $$= 224{,}03$$

   Lakukan perhitungan yang sama sampai data ke-3526.

   Berikut ini adalah beberapa hasil perhitungan jarak Euclidean pada titik pusat klaster 2 :

   $$D(1,2) = \sqrt{(1-0)^2} + (231 - 118)^2 + (3 - 3)^2 + (4 - 4)^2 + (4 - 4)^2 + (27 - 59)^2 + (3 - 4)^2$$
   $$= 117{,}45$$
   $$D(2,2) = \sqrt{(1-0)^2} + (231 - 118)^2 + (4 - 3)^2 + (4 - 4)^2 + (4 - 4)^2 + (28 - 59)^2 + (3 - 4)^2$$
   $$= 117{,}19$$

   Lakukan perhitungan yang sama sampai data ke-3526.

   Berikut ini adalah beberapa hasil perhitungan jarak Euclidean pada titik pusat klaster 3 :

$$D(1,3) = \sqrt{(1-1)^2} + (231 - 29)^2 + (3-4)^2 + (4 - 3)^2 + (4 - 4)^2 + (27 - 22)^2 + (3-6)^2$$

$$= 40840$$

$$D(2,3) = \sqrt{(1-1)^2} + (231 - 29)^2 + (4-4)^2 + (4 - 3)^2 + (4 - 4)^2 + (28 - 22)^2 + (3-6)^2$$

$$= 40850$$

Kemudian dilanjutkan dengan perhitungan yang sama sampai data ke-3526. Berikut ini hasil perhitungan jarak iterasi 1. Berikut ini adalah rekap hasil perhitungan jarak antar data yang dihitung sampai data ke-3526 (Tabel 4).

Tabel 4 Rekap Hasil Perhitungan Jarak Iterasi 1 menggunakan *Euclidean Distance*

No	Jarak C1	Jarak C2	Jarak C3	Klaster
1	224,02	117,45	40840	2
2	224,03	117,19	40850	2
3	116,38	0,00	9297	2
4	0,00	116,41	490	1
5	224,00	118,32	40810	2
….	….	….	….	….
3526	177,77	64,22	24517	2

4. Langkah selanjutnya dari rekap perhitungan jarak data tersebut, kemudian yaitu lakukan perhitungan kembali pusat klaster yang baru atau centroid yang baru dengan menjumlahkan data pada masing-masing atribut berdasarkan hasil perhitungan jarak pada iterasi 1, dilanjut dibagi dengan banyaknya data pada setiap klaster.

Tabel 5 *Centroid* Baru Hasil Iterasi 1

Centroid Baru 1	L/P	DP	K	LRI	SP	U	G
C1	0,6	13,6	4,2	3,7	4,0	25,3	2,8
C2	0,6	155,4	4,3	3,8	4,0	34,1	3,4
C3	1,0	29,1	4,4	3,2	4,0	21,8	6,0

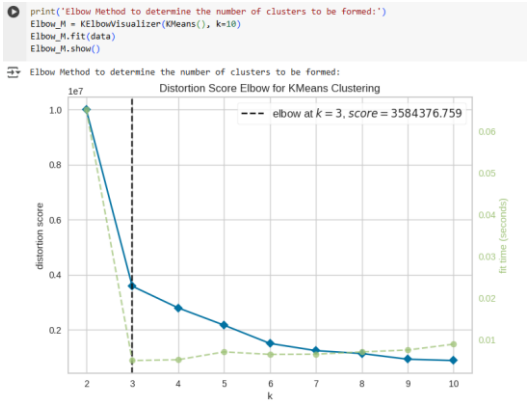
5. Langkah terakhir yaitu mengulangi proses tersebut dari langkah ketiga sampai keempat sampai titik pusat dari setiap klaster tidak berubah dan tidak ada lagi data yang berpindah dari satu klaster ke klaster lainnya. Pada proses selanjutnya, diperoleh pada iterasi kelima titik pusat klaster (centroid) tidak mengalami perubahan kembali dan tidak ditemukan data yang berpindah dari satu klaster ke klaster yang lain. Hal ini menandakan bahwa proses klasterisasi telah mencapai konvergensi, yang berarti setiap data yang telah diklasterkan ke dalam masing-masing klaster telah sesuai dengan karakteristiknya. Berikut ini adalah hasil titik centroid setelah iterasi ke-5 (Tabel 10).

Tabel 6 *Centroid* Baru Hasil Iterasi 5

Centroid Baru 1	L/P	DP	K	LRI	SP	U	G
C1	0,6	22,1	4,2	3,7	4	29	3
C2	0,6	165	4	3,8	4	31	3
C3	1	29,1	4,4	3,176	4	21,8	6

3.2 Implementasi K-Means Clustering menggunakan Python

Implementasi algoritma *K-Means Clustering* dilakukan dengan menggunakan bahasa pemrograman *Python* dalam aplikasi *Google Colab.* Data yang digunakan, diekspor dalam format *excel* ke dalam aplikasi tersebut. Langkah pertama melakukan data cleaning dan data transformation pada data tersebut. Kemudian dilanjutkan menentukan jumlah klaster yang dibutuhkan dalam impelementasi K-Means dalam penelitian ini. Salah satu metode penentuan tersebut dapat menggunakan *elbow method* dimana diperoleh 3 klaster berdasarkan rekomendasi *elbow method* (Gambar 3). Hasil tersebut diperoleh berdasarkan perhitungan nilai *Sum of Square Error* (SSE) dimana nilai SSE tidak lagi mengalami penurunan yang signifikan.



Gambar 3 Hasil Rekomendasi Jumlah Klaster menggunakan *Elbow Method*

Langkah berikutnya yaitu mengelompokkan data tersebut menjadi 3 klaster menggunakan algoritma *K-Means.* Berikut ini implementasinya menggunakan Python.

```
from sklearn.cluster import KMeans
k = 3 #tiga segment / tiga kelompok / 3 cluster
km = KMeans(n_clusters=k, init='random', max_iter=300, random_state=20)
y = km.fit_predict(data)
```

Gambar 4 Penerapan Klasterisasi menggunakan Python

Titik pusat (*centroid*) dari setiap *cluster* yang dihasilkan oleh algoritma *K-Means* ditampilkan menggunakan kode sebagai berikut (Gambar 5). Diperoleh juga hasil perhitungan jarak antar *centroid* menggunakan *Euclidean Distance* menggunakan kode sebagai berikut (Gambar 5).

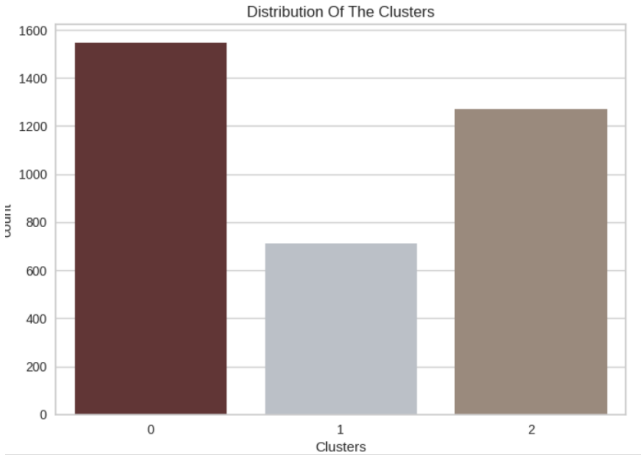
```
from sklearn.metrics import pairwise_distances

centroids = km.cluster_centers_
distances = pairwise_distances(centroids)

print("Centroids:\n", centroids)
print("Jarak antar centroid:\n", distances)
```

```
Centroids:
 [[  0.5631068   13.61294498   4.1987055    3.71003236   3.99029126
   25.35598706   2.81359223]
 [  0.63764045 228.65449438   4.3511236    3.45786517   3.96629213
   28.06039326   3.2738764 ]
 [  0.56816391 113.1284476    4.29708432   4.05200946   3.96296296
   37.37982664   3.46650906]]
Jarak antar centroid:
 [[  0.         215.05926298 100.24201553]
 [215.05926298   0.         115.90305085]
 [100.24201553 115.90305085   0.        ]]
```

Gambar 5 *Centroid* dari Setiap Klaster dan Perhitungan Jarak antar *Centroid*

3.3 Deskripsi Hasil Clustering

Berdasarkan hasil pengolahan data, data pasien BPJS rawat inap RS Islam Assyifa Sukabumi dapat dikelompokkan menjadi 3 klaster. Proses ini menghasilkan klaster 1 berjumlah 1545 pasien, klaster 2 berjumlah 712 pasien, serta klaster 3 berjumlah 1269 pasien. Berikut ini distribusi data dari setiap klaster (Gambar 4).



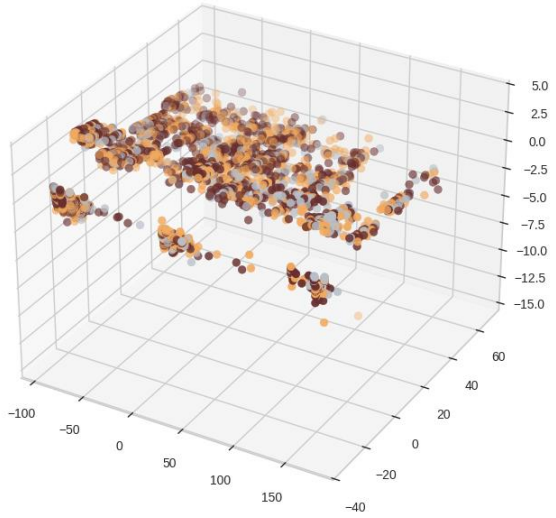
Gambar 6 Distribusi Data setiap Klaster

Klaster 1 merupakan kelompok dengan mayoritas penyakit yang ditemukan adalah kategori penyakit menular yang disebabkan oleh mikroorganisme (WHO, 2010). Penyakit yang paling banyak ditemukan adalah *viral infection, unspecified* sebanyak 483 pasien, diikuti oleh *gastroentritis and colitis of unspecified region* sebanyak 279 kasus dan *bacterial infection, unspecified* sebanyak 177 kasus. Dari total pasien tersebut, didominasi oleh pasien berjenis kelamin perempuan sebanyak 870 pasien. Berdasarkan usia pasien, sebagian besar pasien berasal dari kelompok usia balita (1-5 tahun) dan lansia (46-65 tahun) masing-masing sebanyak 423 pasien dan 290 pasien. Mayoritas pasien dalam kelompok ini menggunakan fasilitas rawat inap sesuai dengan kelasnya yaitu kelas III (510 pasien) dan kelas II (451 pasien). Sebagian besar pasien dalam klaster ini juga mendapatkan pelayanan rawat inap selama 3 hari sebanyak 792 pasien.

Klaster 2 merupakan kelompok dengan mayoritas penyakit yang berkaitan dengan kehamilan, kelahiran, maupun gejala yang harus diidentifikasi melalui pemeriksaan klinikal atau cek laboratorium lanjutan (WHO, 2010). Penyakit yang paling banyak ditemukan adalah *other*

and unspesified abdominal pain sebanyak 75 kasus, diikuti oleh *fetus and newborn affected by other maternal nutrition disorders* sebanyak 73 kasus dan *febrile convulsions* sebanyak 72 kasus. Dari total pasien tersebut, didominasi oleh pasien berjenis kelamin perempuan sebanyak 454 pasien. Berdasarkan usia pasien, kelompok usia pasien terbesar dalam klaster ini berasal dari usia dewasa (26-45 tahun) sebanyak 192 pasien kemudian diikuti usia lansia (46-65 tahun) sebanyak 131 orang. Berdasarkan kelas BPJS, mayoritas pasien dalam klaster memiliki kemiripan dengan klaster 1 yaitu kelas III (296 pasien) dan kelas II (203 pasien). Sebagian besar pasien dalam klaster ini mendapatkan rawat inap selama 3 hari sebanyak 472 pasien.

Klaster 3 merupakan kelompok dengan mayoritas penyakit yang ditemukan adalah penyakit yang berkaitan dengan sistem pernafasan, sistem pencernaan, dan sistem sirkulasi darah (WHO, 2010). Penyakit yang paling banyak ditemukan adalah *bronchitis* sebanyak 220 kasus, diikuti oleh *pneumonia* sebanyak 146 kasus dan *dyspepsia* sebanyak 153 kasus. Dari total pasien tersebut, didominasi oleh pasien berjenis kelamin perempuan sebanyak 721 pasien. Berdasarkan usia pasien, kelompok usia pasien yang terbanyak adalah usia lansia (46-65 tahun) sebanyak 346 pasien diikuti usia dewasa (26-45 tahun) sebanyak 256 pasien. Berdasarkan kelas BPJS, mayoritas pasien dalam klaster ini memiliki kemiripan dengan klaster 1 dan 2 yaitu kelas III (420 pasien) dan kelas II (363 pasien). Berbeda dengan klaster 1, sebagian besar pasien dalam kelompok ini mendapatkan pelayanan rawat inap selama 4 hari dengan jumlah sebanyak 532 pasien.



Gambar 7 Visualisasi Klaster menggunakan *Scatter Plot*

3.4 Evaluasi

Evaluasi *clustering* dalam penelitian ini menggunakan metode *Davies-Bouldin Index* (DBI). Tujuan utama dari evaluasi ini adalah mengukur seberapa baik pemisahan antar klaster dengan mempertimbangkan jarak antar klaster. Pengujian dengan DBI digunakan untuk mengevaluasi seberapa baik hasil *clustering* yang diperoleh. Nilai DBI yang dihasilkan akan menjadi gambaran representatif dari kualitas klaster tersebut, dimana semakin rendah nilai DBI yang diperoleh maka semakin baik kualitas klaster tersebut. Berikut ini merupakan hasil evaluasi DBI yang telah diterapkan menggunakan bahasa pemprograman Python.

```
from sklearn.cluster import KMeans
k = 3 #tiga segment / tiga kelompok / 3 cluster
km = KMeans(n_clusters=k, init='random', max_iter=300, random_state=20)
y = km.fit_predict(data)
```

```
from sklearn.metrics import davies_bouldin_score
dbi = davies_bouldin_score(data, y)
print("Davis-Bouldin Index:", dbi)
```

```
Davis-Bouldin Index: 0.5610335155540453
```

Gambar 8 Penerapan Evaluasi menggunakan Nilai DBI

Tabel 7 Hasil Pengujian Davies-Bouldin Index (DBI)

No	Jumlah Klaster	DBI
1	2	0,631
2	**3**	**0,561**
3	4	0,648
4	5	0,748
5	6	0,706

Berdasarkan hasil pengujian DBI (Tabel 10), didapatkan bahwa nilai k yang optimal adalah k=3 dengan *Davies-Bouldin Index* (DBI) sebesar 0.561. Hal ini menunjukkan bahwa proses *clustering* yang diterapkan dengan jumlah sebanyak 3 klaster memiliki nilai DBI yang lebih rendah dibandingkan dengan proses *clustering* dengan jumlah klaster kurang dari 3 atau lebih dari 3. Hasil ini selaras dengan hasil penelitian terdahulu yang menunjukkan pengelompokkan pasien menjadi 3 klaster memiliki kualitas yang terbaik (Herdiaman et al., 2024).

4 Kesimpulan

Pengelompokkan data rekam medis pasien BPJS rawat inap pada periode triwulan keempat 2024 menggunakan analisis *K-Means Clustering* diperoleh hasil yang optimal, dimana diperoleh 3526 data yang terbagi dalam 3 klaster dan nilai Davies-Bouldin Index (DBI) sebesar 0,561 atau mendekati 0.

Klaster 1 merupakan kelompok penyakit menular yang disebabkan oleh mikroorganisme dengan jumlah 1545 data dan usia pasien terbanyak berasal dari usia balita (1-5 tahun), terdiri dari 675 pasien laki-laki dan 870 pasien perempuan dengan penyakit *viral infection, unspecified* yang paling umum. Jumlah lama rawat inap terbanyak pada 3 hari dan jumlah kelas BPJS yang digunakan terbanyak pada kelas III.

Klaster 2 merupakan kelompok penyakit yang berkaitan dengan masalah kehamilan, kelahiran maupun gejala yang harus diidentifikasi melalui pemeriksaan klinikal atau laboratorium lanjutan dengan jumlah 712 data dan usia pasien terbanyak berasal dari usia dewasa (26-45 tahun), terdiri dari 258 pasien laki-laki dan 454 pasien perempuan dengan penyakit *other and unspesified abdominal pain* yang paling umum. Jumlah lama rawat inap terbanyak pada 3 hari dan jumlah kelas BPJS yang digunakan terbanyak pada kelas III.

Klaster 3 merupakan kelompok penyakit yang berkaitan dengan sistem pernafasan, sistem pencernaan, dan sistem sirkulasi darah dengan jumlah 1269 data dan usia pasien terbanyak berasal

dari usia lansia (46-55 tahun) dengan penyakit *bronchitis* yang paling umum. Jumlah lama rawat inap terbanyak pada 4 hari dan jumlah kelas BPJS yang digunakan terbanyak pada kelas III.

Berdasarkan hasil penelitian ini dapat dijadikan acuan bagi rumah sakit dalam merencanakan alokasi sumber daya rumah sakit, terutama penambahan jumlah kamar rawat inap yang sesuai dengan kelasnya. Saran untuk penelitian berikutnya adalah penggunaan teknik *data mining* lainnya sebagai bentuk upaya optimalisasi manajemen rumah sakit.

Daftar Pustaka

Ali, A., & Masyfufah, L. (2020). Klasterisasi Pasien BPJS Dengan Metode K-Means Clustering Guna Menunjang Program Jaminan Kesehatan Nasional Di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *Jurnal Wiyata*, *8*(1), 8–22. https://doi.org/p://dx.doi.org/10.56710/wiyata.v8i1.427

BPJS Kesehatan. (2024). *Laporan Pengelolaan Program Tahun 2023 & Laporan Keuangan Tahun 2023 (Auditan)*.

Firmansah, D. A., Rohman, R. S., & Ermawati, E. (2021). Penerapan Metode Ward and Peppard - Cassidy Pada Perencanaan Strategis Sistem Informasi Rumah Sakit Islam Assyifa Sukabumi. *Simpatik: Jurnal Sistem Informasi Dan Informatika*, *1*(1), 43–52. https://doi.org/10.31294/simpatik.v1i1.407

Herdiaman, E. A., Sudiarjo, A., Hikmatyar, M., Informatika, T., Perjuangan, U., & Barat, J. (2024). Klasterisasi Pasien pada RSUD Ciamis Menggunakan Metode K-Means. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, *12*(3), 3558–3546. https://doi.org/http://dx.doi.org/10.23960/jitet.v12i3S1.5124

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Ningsih, Y. G., Samosir, K., & Satria, B. (2025). Segmentasi Pasien Rumah Sakit Berdasarkan Pola Kunjungan Menggunakan Algoritma K-Means Clustering untuk Optimasi Layanan Medis. *Jurnal Pendidikan Sains Dan Komputer*, *5*(1), 59–69. https://doi.org/https://doi.org/10.47709/jpsk.v5i01.5492 Segmentasi

Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022 Tentang Rekam Medis, Kementerian Kesehatan Republik Indonesia 1 (2022).

Saputra, I. (2023). *Belajar Mudah Data Mining untuk Pemula* (Cetakan 1). Informatika Bandung.

Sulastri, H., & Gufroni, A. I. (2017). Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *3*(2), 299–305. https://doi.org/10.25077/teknosi.v3i2.2017.299-305

WHO. (2010). International Statistical Classification of Diseases 10th Revision, Version for 2010. In *World Health Organization* (Vols. 1 & 3). https://www.who.int/publications/m/item/icd-10-updates-2010