

## Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means

Jeri Wandana<sup>1✉</sup>, Sarjon Defit<sup>2</sup>, Sumijan<sup>3</sup>

<sup>1,2,3</sup>Universitas Putra Indonesia YPTK Padang  
[jwandana@gmail.com](mailto:jwandana@gmail.com)

### Abstract

Patient histories who use the services of Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan are stored in medical record data. Each medical record data contains important information that is very valuable and can be processed to explore new knowledge using a data mining approach. This study aims to help Prof. Dr. Tabrani hospital in classifying patient data who use BPJS Kesehatan, so that the pattern of disease spread is known based on class of service. The data used is patient medical record data in 2019 from October to December, the data will be processed using the K-Means Clustering algorithm with a total of 3 clusters. In cluster 0 (H0) there are 3 patients who are dominated by A09.9 disease (Diarrhea / Dysentery) in Class 2 and Class 3, for cluster 1 (H1) there are 5 patients with more diverse types of disease, while for cluster 2 (H2) there are 5 patients who are predominantly K30 disease (Dyspepsia) in Class 1.

Keywords: Medical Record, BPJS, Data Mining, K-Means, Cluster.

### Abstrak

Riwayat pasien yang menggunakan layanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan tersimpan pada data rekam medis. Setiap data rekam medis berisikan informasi penting yang sangat berharga dan dapat diolah untuk menggali pengetahuan baru menggunakan pendekatan *data mining*. Penelitian ini bertujuan untuk membantu rumah sakit Prof. Dr. Tabrani dalam mengelompokkan data pasien pengguna BPJS Kesehatan, sehingga diketahui pola penyebaran penyakit berdasarkan kelas layanan. Data yang digunakan adalah data rekam medis pasien tahun 2019 periode bulan Oktober sampai dengan Desember, data tersebut akan diproses menggunakan algoritma *K-Means Clustering* dengan jumlah *cluster* sebanyak 3. Pada *cluster* 0 (H0) terdapat 3 pasien yang didominasi penyakit A09.9 (Diare/Disentri) pada Kelas 2 dan Kelas 3, untuk *cluster* 1 (H1) terdapat 5 pasien dengan jenis penyakit yang lebih beragam, sedangkan untuk *cluster* 2 (H2) terdapat 5 pasien yang didominasi penyakit K30 (*Dyspepsia*) pada Kelas 1.

Kata kunci: Rekam Medis, BPJS, *Data Mining*, *K-Means*, *Cluster*.

© 2020 JIDT

### 1. Pendahuluan

Rumah sakit Prof. Dr. Tabrani merupakan salah satu lembaga penyedia jasa layanan kesehatan yang berdiri sejak tahun 1977 di Kota Pekanbaru dan terus membuktikan eksistensinya dengan tetap menjadi pilihan berobat bagi masyarakat.

Dengan adanya fasilitas jaminan sosial BPJS Kesehatan yang diselenggarakan pemerintah dengan menjunjung prinsip gotong royong dalam penyetaraan jaminan kesehatan masyarakat, banyak pasien yang menggunakan fasilitas tersebut. Rumah sakit Prof. Dr. Tabrani 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 jenis kelas BPJS Kesehatan pasien.

Ekstraksi informasi yang berharga pada data rekam medis dapat dilakukan dengan menggunakan metode *Clustering* berdasarkan atribut kode penyakit, umur, jenis kelamin, jenis BPJS, *referred*, kelas BPJS, dan faskes.

Penelitian yang pernah dilakukan sebelumnya menggunakan algoritma *K-Means Clustering* diantaranya yaitu Sembiring dkk (2020) dalam pengklasteran daerah pungutan liar di Kabupaten Sukabumi, *cluster* 1 tingkat pungutan liar tinggi yaitu pada daerah Prabaya, *cluster* 2 tingkat pungutan sedang yaitu pada 35 daerah, *cluster* 3 tingkat pungutan rendah yaitu pada 43 daerah [1]. Penelitian Purnama dkk (2019) algoritma *K-Means Clustering* dalam pemetaan jumlah tindak pidana *illegal fishing*, *cluster* 1 berjumlah 6 kasus sebagai peringkat 3, *cluster* 2 berjumlah 2 kasus sebagai peringkat 4, *cluster* 3 berjumlah 17 kasus sebagai peringkat 2, dan *cluster* 4 berjumlah 242 kasus sebagai peringkat 1 [2]. Penelitian Ghofar dkk (2018) dalam klasterisasi pelanggan UMS Store, pelanggan potensial tinggi berada pada *cluster* 3, pelanggan

berpotensi menengah berada pada *cluster* 2 dan pelanggan kurang berpotensi berada pada *cluster* 1 [3].

Penelitian Bastian dkk (2018) dalam klasterisasi pada penyakit menular manusia di Kabupaten Majalengka, hasil penelitian dapat digunakan untuk pengendalian persediaan obat pada penyakit menular serta dapat dilihat daftar Puskesmas yang banyak menderita penyakit menular [4]. Penelitian Maulida (2018) pada penerapan *data mining* dalam mengelompokkan kunjungan wisatawan ke objek wisata unggulan di Provinsi DKI Jakarta, *cluster* 1 kunjungan tertinggi dengan nilai *centroid* 15.438.488, *cluster* 2 kunjungan sedang dengan nilai *centroid* 4.464.577, dan *cluster* 3 kunjungan rendah dengan nilai *centroid* 342.332 [5].

Penelitian Asroni dkk (2018) dalam pengelompokkan data calon mahasiswa baru menggunakan algoritma *K-Means Clustering*, penelitian menunjukkan bahwa *K-Means* berhasil dalam pengelompokkan data dengan hasil bahwa jurusan Pendidikan Dokter merupakan yang paling favorit dengan persentase 33% [6]. Penelitian Umar dkk (2018) dalam identifikasi keaslian uang kertas Rupiah menggunakan algoritma *K-Means*, akurasi dalam pengelompokkan pengenalan citra pada permukaan uang kertas mencapai 96,67% [7].

Penelitian Sukamto dkk (2018) dalam penentuan daerah rawan titik api, *cluster* dibagi kedalam 3 kelompok, yaitu 133 daerah sangat rawan titik api, 101 daerah rawan titik api, dan 77 daerah tidak rawan titik api [8]. Penelitian Aprianti dkk (2018) dalam klasterisasi data kecelakaan lalu lintas jalan raya, hasil klasterisasi didapatkan pengetahuan bahwa salah faktor kecelakaan adalah pengemudi, dengan korban didominasi oleh usia 9 sampai dengan 28 tahun [9].

Penelitian Wiza (2019) dalam klasterisasi kekerasan seksual terhadap anak, *cluster* 1 didominasi oleh kriteria kekerasan pencabulan pada anak usia 3 sampai dengan 16 tahun dengan faktor pemicu adalah adanya kesempatan [10]. Penelitian Puntoriza (2020) dalam menganalisa persebaran UMKM Kota Malang menggunakan *K-Means Clustering*, ditemukan bahwa Kecamatan Belimbing masuk kedalam *cluster* 1 dengan potensi tinggi, Kecamatan Klojen ke dalam *cluster* 2 dengan potensi sedang, dan Kecamatan Kedungkandang, Sukun, serta Kecamatan Lowokwaru ke dalam *cluster* 3 dengan potensi rendah [11].

Penelitian Am dkk (2019) dalam penentuan mutu kelapa sawit menggunakan algoritma *K-Means Clustering*, hasil penelitian diperoleh *cluster* 0 dengan kategori baik sebanyak 12 data, *cluster* 1 dengan kategori sangat baik sebanyak 13 data dan *cluster* 2 dengan kategori kurang baik sebanyak 6 data [12]. Penelitian Salam dkk (2020) dalam pengklasteran rekomendasi penerima beasiswa PPA di UDINUS, *dataset* berjumlah 441 pada tahun 2016, dari hasil perhitungan terdapat 154 mahasiswa yang direkomendasikan mendapat beasiswa PPA, 287 lainnya tidak direkomendasikan [13].

Penelitian Sarasvananda dkk (2019) dalam upaya memperkirakan biaya rawat inap pasien menggunakan algoritma *K-Means* dengan *Semantic Similarity*, kombinasi algoritma *K-Means* dan *Semantic Similarity* memberikan akurasi sebesar 91,78% dalam memperkirakan biaya rawat inap pasien [14]. Penelitian Annur (2019) 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 [15].

## 2. Metodologi Penelitian

Metodologi penelitian dilakukan secara sistematis mulai dari pengumpulan data yang bersumber dari rumah sakit Prof. Dr. Tabrani Pekanbaru tahun 2019 periode bulan Oktober sampai dengan Desember dan sumber jurnal yang relevan dengan masalah yang diteliti.

*Knowledge Discovery in Database* (KDD) adalah cara *non-trivial* dalam upaya menemukan pola pada sebuah data yang bersifat baru, dapat dimengerti dan bermanfaat. Pada tahapan KDD terdapat proses *data mining*, yaitu melakukan analisa dari sejumlah besar kumpulan data observasi.

*Clustering* adalah suatu alat bantu pada *data mining* yang bertujuan mengelompokkan objek-objek kedalam *cluster-cluster*. *Cluster* adalah sekelompok atau sekumpulan objek-objek data yang mempunyai kemiripan karakteristik satu sama lain dalam *cluster* yang sama dan berbeda karakteristik terhadap objek-objek yang berbeda *cluster* [16]. Diantara beberapa teknik *clustering* pada *data mining* yang paling populer adalah *K-Means Clustering*. Metode *K-Means Clustering* menjadi sangat populer dikarenakan dapat dengan cepat dan efisien mengelompokkan sejumlah besar data dengan pemodelan tanpa supervisi (*unsupervised learning*) [17]. Diantara keunggulan yang dimiliki algoritma *K-Means Clustering* adalah [18]:

- Relatif mudah diterapkan.
- Menskalakan ke kumpulan data besar.
- Menjamin konvergensi.
- Dapat memperbaharui posisi *centroid*.
- Mudah beradaptasi.
- Menggeneralisasi *cluster* dengan berbagai bentuk dan ukuran, seperti *cluster elips*.

Tahapan pemrosesan menggunakan algoritma *K-Means Clustering* dilakukan dengan beberapa langkah.

- Menentukan jumlah *cluster* yang diinisialkan dengan K;
- Menentukan *centroid* awal dengan cara *random*;
- Menentukan *centroid* terdekat dari setiap titik data dengan menghitung jarak ke setiap *centroid* menggunakan rumus *Euclidean Distance*, seperti pada persamaan (1).

$$Q_{(j,r)} = \sqrt{(m_{1j} - m_{1r})^2 + (m_{2j} - m_{2r})^2 + \dots + (m_{nj} - m_{nr})^2} \quad (1)$$

Dimana :

$Q_{(j,r)}$  = Jarak data ke- $j$  ke pusat  $cluster$   $r$

$m_{nj}$  = Data ke- $j$  pada atribut data ke- $n$

$m_{nr}$  = Titik pusat ke- $r$  pada atribut ke- $n$

d. Mengelompokkan data dengan jarak terdekat

e. Menentukan  $centroid$  baru dengan menghitung nilai median untuk setiap  $cluster$ , seperti pada persamaan (2).

$$S_l = \frac{1}{Z_l} (T_{1l} + T_{2l} + \dots + T_{nl}) \quad (2)$$

Dimana :

$S_l$  = Rata-rata  $cluster$  baru

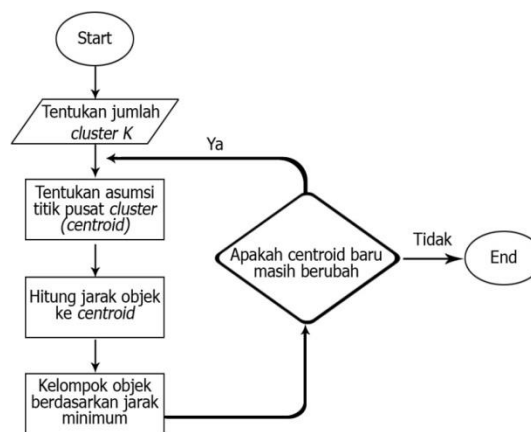
$Z_l$  = Jumlah percobaan pola pada  $cluster$  ke- $l$

$T_{nl}$  = Pola ke- $n$  yang menjadi bagian dari  $cluster$  ke- $l$

f. Kembali ke langkah 3

g. Proses berhenti jika tidak ada perubahan data dalam penetapan  $cluster$

Langkah-langkah tersebut digambarkan dalam bentuk *flowchart* seperti pada Gambar 1.



Gambar 1. *Flowchart* Algoritma  $K$ -Means

### 3. Hasil dan Pembahasan

Berdasarkan penelitian yang dilakukan, hasil dan pembahasan adalah seperti berikut:

- Input dataset* rekam medis pasien BPJS Kesehatan rumah sakit Prof. Dr. Tabrani Pekanbaru tahun 2019 periode bulan Oktober sampai dengan Desember. Sampel data rekam medis pasien disajikan seperti pada Tabel 1, untuk kolom nama pasien akan diinisialkan dengan kode PSN1 sampai dengan PSN13.
- Transformasi data dengan bentuk kualitatif menjadi data kuantitatif, yaitu kode penyakit, umur, jenis kelamin, jenis BPJS, *referred*, kelas BPJS dan faskes. Hasil transformasi seperti pada Tabel 2.
- Tentukan jumlah  $cluster$ , pada penelitian ini digunakan 3  $cluster$ , yaitu H0, H1 dan H2. Dimana H0 mewakili  $Cluster$  0, H1 mewakili  $Cluster$  1, dan H2 =  $Cluster$  2.

Tabel 1. Sample Data Rekam Medis Pasien

No	Nama	Kode Penyakit	Umur (Tahun)	Jenis Kelamin	Jenis BPJS	Referred	Kelas BPJS	Faskes
1	PSN1	A09.9	9	Laki-laki	Karyawan swasta	IGD	Kelas 2	Siak Hulu 1
2	PSN2	I84.2	62	Perempuan	Mandiri	POLI	Kelas 3	BP.Annisa
3	PSN3	Z53.2	27	Laki-laki	Karyawan Swasta	IGD	Kelas 2	Šari Husada
4	PSN4	A90	39	Perempuan	Karyawan Swasta	IGD	Kelas 2	K. Salsa
5	PSN5	J40	54	Perempuan	PBI	IGD	Kelas 3	Harapan Raya
6	PSN6	K30	59	Perempuan	Mandiri	IGD	Kelas 2	Lima Puluh
7	PSN7	M06.9	50	Perempuan	PBI	POLI	Kelas 3	simpang tiga
8	PSN8	A09.9	37	Laki-laki	Mandiri	IGD	Kelas 3	Nayaka Husada
9	PSN9	K30	55	Perempuan	PNS	POLI	Kelas 1	Simpang Tiga
10	PSN10	B34.9	30	Laki-laki	Karyawan	IGD	Kelas 2	Yahya Medika
11	PSN11	N80.0	47	Perempuan	Mandiri	POLI	Kelas 3	Garuda
12	PSN12	K30	64	Perempuan	Mandiri	IGD	Kelas 1	Pratama Kartika
13	PSN13	K30	29	Perempuan	Mandiri	IGD	Kelas 3	Annisa Medika

Dimana Tabel 1 merupakan sampel data rekam medis pasien yang berjumlah 50 data.

Tabel 2. Transformasi Sample Data Rekam Medis Pasien

No	Nama	Kode Penyakit	Umur	Jenis Kelamin	Jenis BPJS	Referred	Kelas BPJS	Faskes
1	PSN1	102	0	1	6	1	2	129
2	PSN2	135	2	0	1	2	3	13
3	PSN3	195	1	1	6	1	2	125
4	PSN4	104	1	0	6	1	2	59
5	PSN5	147	2	0	1	1	1	112
6	PSN6	140	2	0	5	1	3	38
7	PSN7	147	2	0	1	1	2	92
8	PSN8	163	1	0	5	2	3	136
9	PSN9	102	1	1	1	1	3	99
10	PSN10	147	2	0	2	2	1	136
11	PSN11	147	1	0	1	1	3	47
12	PSN12	106	1	1	6	1	2	147
13	PSN13	178	1	0	1	2	3	35

Dimana pada Tabel 2 terlihat hasil transformasi data kedalam bentuk numerik.

d. Lakukan proses iterasi pertama dengan menggunakan rumus korelasi antara tiga objek (*Euclidean Distance*) agar dapat diketahui jarak tiap data ke pusta *centroid*, *centroid* awal untuk H0 diambil dari data urutan 1, H1 diambil dari data urutan 11 dan H2 diambil dari data urutan 8, seperti pada Tabel 3.

Tabel 3. *Centroid* Awal

<b>Centroid Iterasi 1</b>								
H0	102	0	1	6	1	2	129	
H1	147	1	0	1	1	3	47	
H2	163	1	0	5	2	3	136	

Dimana pada Tabel 3 dapat dilihat nilai *centroid* awal untuk proses iterasi 1.

a. Hitung jarak atribut ke *centroid* 0, perhitungan diambil dari data pada urutan 1

$$T_0 = \sqrt{\frac{(102-102)^2+(0-0)^2+(1-1)^2+(6-6)^2+(1-1)^2+(2-2)^2+(129-129)^2}{(1-1)^2+(2-2)^2+(129-129)^2}}$$

$$= 0$$

b. Hitung jarak atribut ke *centroid* 1, perhitungan diambil dari data pada urutan 1

$$T_1 = \sqrt{\frac{(102-147)^2+(0-1)^2+(1-0)^2+(6-1)^2+(1-1)^2+(2-3)^2+(129-47)^2}{(1-1)^2+(2-3)^2+(129-47)^2}}$$

$$= 93,7$$

c. Hitung jarak atribut ke *centroid* 2, perhitungan diambil dari data pada urutan 1

$$T_2 = \sqrt{\frac{(102-163)^2+(0-1)^2+(1-0)^2+(6-5)^2+(1-2)^2+(2-3)^2+(129-136)^2}{(1-2)^2+(2-3)^2+(129-136)^2}}$$

$$= 61$$

d. Kelompokkan tiap data dengan jarak terdekat ke pusat *cluster*. Hasil perhitungan iterasi pertama dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pengelompokkan Sample Data Iterasi 1

No	Nama	H0	H1	H2	Jarak Terdekat	H0	H1	H2
1	PSN1	0	93,7	61	0	1	0	0
2	PSN2	120,7	36,1	126	36,1	0	1	0
3	PSN3	93,1	91,7	34	33,9	0	0	1
4	PSN4	70,0	44,9	97	44,9	0	1	0
5	PSN5	48,4	65,0	29	29,2	0	0	1
6	PSN6	98,7	12,1	101	12,1	0	1	0
7	PSN7	58,5	45,0	47	45,0	0	1	0
8	PSN8	61,4	90,5	0	0,0	0	0	1
9	PSN9	30,4	68,8	71	30,4	1	0	0
10	PSN10	45,8	89,0	16	16,4	0	0	1
11	PSN11	93,7	0,0	91	0,0	0	1	0
12	PSN12	18,5	108,2	58	18,5	1	0	0
13	PSN13	121	33,3	102	33,3	0	1	0
Jumlah						3	6	4

Dimana Tabel 4 merupakan hasil pengelompokkan iterasi 1.

Berdasarkan hasil iterasi pertama, maka diperoleh anggota pada H0 berjumlah 3 data, anggota pada H1 berjumlah 6 data, serta anggota pada H3 berjumlah 4 data.

e. Penentuan *centroid* baru dengan perhitungan nilai median setiap anggota *cluster* untuk iterasi kedua

a. Perhitungan *centroid* 0 baru untuk iterasi kedua

$$H0_{(S1)} = \frac{(102+102+106)}{3} = 103,3$$

$$H0_{(S2)} = \frac{0+1+1}{3} = 0,7$$

$$H0_{(S3)} = \frac{(1+1+1)}{3} = 1$$

$$H0_{(S4)} = \frac{(6+1+6)}{3} = 4,3$$

$$H0_{(S5)} = \frac{1+1+1}{3} = 1$$

$$H0_{(S6)} = \frac{(2+1+1)}{3} = 2,3$$

$$H0_{(S7)} = \frac{(129+99+147)}{3} = 125$$

b. Perhitungan *centroid* 1 baru untuk iterasi kedua

$$H1_{(S1)} = \frac{(135+104+140+147+147+178)}{6} = 141,8$$

$$H1_{(S2)} = \frac{(2+1+2+2+1+1)}{6} = 1,5$$

$$H1_{(S3)} = \frac{(0+0+0+0+0+0)}{6} = 0$$

$$H1_{(S4)} = \frac{(1+6+5+1+1+1)}{6} = 2,5$$

$$H1_{(S5)} = \frac{2+1+1+1+1+2}{6} = 2,5$$

$$H1_{(S6)} = \frac{(3+2+3+2+3+3)}{6} = 2,7$$

$$H1_{(S7)} = \frac{(13+59+38+92+47+35)}{6} = 47,3$$

c. Perhitungan *centroid* 2 baru untuk iterasi kedua

$$H2_{(S1)} = \frac{(195+147+163+147)}{4} = 163$$

$$H2_{(S2)} = \frac{(1+2+1+2)}{4} = 1,5$$

$$H2_{(S3)} = \frac{(1+0+0+0)}{4} = 0,25$$

$$H2_{(S4)} = \frac{(6+1+5+2)}{4} = 3,5$$

$$H2_{(S5)} = \frac{(1+1+2+2)}{4} = 1,5$$

$$H2_{(S6)} = \frac{(2+1+3+1)}{4} = 1,75$$

$$H2_{(S7)} = \frac{(125+112+136+136)}{4} = 127,25$$

Hasil perhitungan *centroid* baru untuk iterasi kedua dapat dilihat pada Tabel 5.

Tabel 5. *Centroid* Baru Iterasi 2

<i>Centroid</i> Iterasi 2							
H0	103,3	0,7	1,0	4,3	1,0	2,3	125,0
H1	141,8	1,5	0,0	2,5	2,5	2,7	47,3
H2	163,0	1,5	0,3	3,5	1,5	1,8	127,3

Dimana Tabel 5 merupakan *centroid* baru untuk iterasi 2

f. Selanjutnya lakukan kembali perhitungan untuk iterasi kedua menggunakan *centroid* baru.

a. Hitung jarak atribut ke *centroid* 0, perhitungan diambil dari data pada urutan 1

$$T_0 = \sqrt{(102-103,3)^2 + (0-0,7)^2 + (1-1)^2 + (6-4,3)^2 + (1-1)^2 + (2-2,3)^2 + (129-125)^2}$$

$$= 4,6$$

b. Hitung jarak atribut ke *centroid* 1, perhitungan diambil dari data pada urutan 1

$$T_1 = \sqrt{(102-140,8)^2 + (0-1,4)^2 + (1-0)^2 + (6-2,8)^2 + (1-2,8)^2 + (2-2,8)^2 + (129-38,5)^2}$$

$$= 98,6$$

c. Hitung jarak atribut ke *centroid* 2, perhitungan diambil dari data pada urutan 1

$$T_2 = \sqrt{(102-159,8)^2 + (0-1,6)^2 + (1-0,2)^2 + (6-3)^2 + (1-1,4)^2 + (2-1,8)^2 + (129-120,2)^2}$$

$$= 58,6$$

d. Kelompokkan tiap data dengan jarak terdekat ke pusat *cluster*. Hasil perhitungan iterasi kedua dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengelompokkan Iterasi 2

No	Nama	H0	H1	H2	Jarak Terdekat	H0	H1	H2
1	PSN1	4,6	91,0	61	4,6	1	0	0
2	PSN2	116,	35,0	118	35,0	0	1	0
3	PSN3	91,7	94,2	32	32,2	0	0	1
4	PSN4	66,0	39,8	90	39,8	0	1	0
5	PSN5	45,7	64,9	22	22,3	0	0	1
6	PSN6	94,4	10,0	92	10,0	0	1	0
7	PSN7	54,9	45,0	39	38,8	0	0	1
8	PSN8	60,7	91,2	9	9,0	0	0	1
9	PSN9	26,3	65,3	67	26,3	1	0	0
10	PSN10	45,2	88,8	18	18,3	0	0	1
11	PSN11	89,5	5,6	82	5,6	0	1	0
12	PSN12	22,2	106,0	60	22,2	1	0	0
13	PSN13	117,	38,2	94	38,2	0	1	0
Jumlah						3	5	5

Dimana Tabel 6 merupakan hasil pengelompokkan iterasi 2.

Hasil iterasi kedua diperoleh anggota H0 berjumlah 5 data, H1 berjumlah 3 data, dan H2 berjumlah 5 data.

g. Penentuan *centroid* baru dengan perhitungan nilai median setiap anggota *cluster* untuk iterasi ketiga

a. Perhitungan *centroid* 0 baru untuk iterasi ketiga

$$H0_{(S1)} = \frac{(102+102+106)}{3} = 103,3$$

$$H0_{(S2)} = \frac{(0+1+1)}{3} = 0,7$$

$$H0_{(S3)} = \frac{(1+1+1)}{3} = 1$$

$$H0_{(S4)} = \frac{(6+1+6)}{3} = 4,3$$

$$H0_{(S5)} = \frac{(1+1+1)}{3} = 1$$

$$H0_{(S6)} = \frac{(2+3+2)}{3} = 2,3$$

$$H0_{(S7)} = \frac{(129+99+147)}{3} = 125$$

b. Perhitungan *centroid* 1 baru untuk iterasi ketiga

$$H1_{(S1)} = \frac{(135+104+140+147+178)}{5} = 140,8$$

$$H1_{(S2)} = \frac{(2+1+2+1+1)}{5} = 1,4$$

$$H1_{(S3)} = \frac{(0+0+0+0+0)}{5} = 0$$

$$H1_{(S4)} = \frac{(1+6+5+1+1)}{5} = 2,8$$

$$H1_{(S5)} = \frac{(2+1+1+1+2)}{5} = 2,8$$

$$H1_{(S6)} = \frac{(3+2+3+3+3)}{5} = 2,8$$

$$H1_{(S7)} = \frac{(13+59+38+47+35)}{5} = 38,4$$

c. Perhitungan *centroid* 2 baru untuk iterasi ketiga yang berjumlah 5 data pasien

$$H2_{(S1)} = \frac{(195+147+147+163+147)}{5} = 159,8$$

$$H2_{(S2)} = \frac{(1+2+2+1+2)}{5} = 1,6$$

$$H2_{(S3)} = \frac{(1+0+0+0+0)}{5} = 0,2$$

$$H2_{(S4)} = \frac{(6+1+1+5+2)}{5} = 3$$

$$H2_{(S5)} = \frac{(1+1+1+2+2)}{5} = 1,4$$

$$H2_{(S6)} = \frac{(2+1+2+3+1)}{5} = 1,8$$

$$H2_{(S7)} = \frac{(125+112+92+136+136)}{5} = 120,2$$

Hasil perhitungan *centroid* baru untuk iterasi ketiga dapat dilihat pada Tabel 7.

Tabel 7. *Centroid* Baru Iterasi 3

<i>Centroid</i> Iterasi 3							
H0	103,3	0,7	1,0	4,3	1,0	2,3	125,0
H1	140,8	1,4	0,0	2,8	2,8	2,8	38,4
H2	159,8	1,6	0,2	3,0	1,4	1,8	120,2

Dimana Tabel 7 merupakan *centroid* baru untuk iterasi 3.

h. Lakukan perhitungan iterasi ketiga menggunakan *centroid* baru.

a. Hitung jarak atribut ke *centroid* 0, perhitungan diambil dari data pada urutan 1

$$T_0 = \sqrt{(102-103,3)^2 + (0-0,7)^2 + (1-1)^2 + (6-4,3)^2 + (1-1)^2 + (2-2,3)^2 + (129-125)^2}$$

$$= 4,6$$

b. Hitung jarak atribut ke *centroid* 1, perhitungan diambil dari data pada urutan 1

$$T_1 = \sqrt{(102-195)^2 + (2-1)^2 + (1-1)^2 + (1-6)^2 + (1-1)^2 + (1-2)^2 + (7-125)^2}$$

$$= 150,33$$

c. Hitung jarak atribut ke *centroid* 2, perhitungan diambil dari data pada urutan 1

$$T_2 = \sqrt{(148-195)^2 + (1-1)^2 + (0-1)^2 + (5-6)^2 + (1-1)^2 + (3-2)^2 + (35-125)^2}$$

$$= 101,55$$

d. Kelompokkan tiap data dengan jarak terdekat ke pusat *cluster*. Hasil perhitungan iterasi ketiga dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pengelompokkan Iterasi 3

No	Nama	H0	H1	H2	Jarak Terdekat			
					H0	H1	H2	
1	PSN1	4,6	98,6	58,6	4,6	1	0	0
2	PSN2	116,	26,1	110,1	26,1	0	1	0
3	PSN3	91,7	102,2	35,7	35,7	0	0	1
4	PSN4	66,0	42,3	82,9	42,3	0	1	0
5	PSN5	45,7	73,9	15,4	15,4	0	0	1
6	PSN6	94,4	3,0	84,6	3,0	0	1	0
7	PSN7	54,9	54,0	31,0	31,0	0	0	1
8	PSN8	60,7	100,1	16,3	16,3	0	0	1
9	PSN9	26,3	72,0	61,6	26,3	1	0	0
10	PSN10	45,2	97,8	20,4	20,4	0	0	1
11	PSN11	89,5	10,9	74,4	10,9	0	1	0
12	PSN12	22,2	114,1	60,2	22,2	1	0	0
13	PSN13	117,	37,4	87,2	37,4	0	1	0
Jumlah					3	5	5	

Pada Tabel 8 dapat dilihat bahwa pada iterasi ketiga memiliki kesamaan data dengan anggota pada iterasi kedua, maka proses iterasi dapat dihentikan.

#### 4. Kesimpulan

Berdasarkan hasil *clustering* yang dilakukan, dapat disimpulkan bahwa proses *clustering* menggunakan algoritma *K-Means Clustering* telah berhasil melakukan pengelompokkan data rekam medis pasien BPJS ke dalam 3 *cluster* dengan kriteria yang berbeda. Ditemukan pengetahuan bahwa penyebaran jenis penyakit dengan kode A09.9 (Diare/Disentri) mendominasi pada *cluster* 0, dan penyebaran jenis penyakit yang lebih beragam pada *cluster* 1, sedangkan

penyebaran jenis penyakit dengan kode K30 (*Dyspepsia*) lebih didominasi pada *cluster* 2.

## Daftar Rujukan

- [1] Sembiring, F., Octaviana, O., & Saepudin, S. (2020). Implementasi Metode K-Means dalam Pengklasteran Daerah Pungutan Liar di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan dan Pencatatan Sipil). *Jurnal Tekno Insentif*, 14(1), 40-47. DOI: <https://doi.org/10.36787/jti.v14i1.165> .
- [2] Purnama, J. J, Nurfalah, R., Rahayu, S., & Novitasari, H. B. (2019). Analisa Algoritma K-Means Clustering Pemetaan Jumlah Tindak Pidana. *KLIK (Kumpulan Jurnal Ilmiah Komputer)*, 6(2), 128-142. DOI: <http://dx.doi.org/10.20527/klik.v6i2.208> .
- [3] Ghofar, M. A., & Kurniawan, Y. I. (2018). Aplikasi Pengelompokan Pelanggan Pada UMS Store Menggunakan Algoritma K-Means. *Jurnal Teknologi dan Manajemen Informatika*, 4(1), 196-203. DOI: <https://doi.org/10.26905/jtmi.v4i1.1772> .
- [4] Bastian, A. (2018). Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi (Journal of Information System)*, 14(1), 28-34.
- [5] Maulida, L. (2018). Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167-174. DOI: <http://dx.doi.org/10.14421/jiska.2018.23-06> .
- [6] Asroni, Fitri, H., Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus : Fakultas Kedokteran dan Ilmu Kesehatan dan Fakultas Ilmu Sosial dan Ilmu Politik). *Jurnal Ilmiah Semesta Teknika*, 21(1), 60-64. DOI: <https://doi.org/10.18196/st.211211> .
- [7] Umar, R., Riadi, I., Miladiah, M. (2018). Sistem Identifikasi Keaslian Uang Kertas Rupiah Menggunakan Metode K-Means Clustering. *Jurnal Teknologi Informasi*, 17(2), 179-185. DOI: <https://doi.org/10.33633/tc.v17i2.1681> .
- [8] Sukamto, S., Id, I. D., & Angraini, T. R. (2018). Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means. *JUITA (Jurnal Informatika)*, 6(2), 137-147. DOI: <http://dx.doi.org/10.30595/juita.v6i2.3172> .
- [9] Aprianti, W., & Permadi, J. (2018). *K-Means Clustering* untuk Data Kecelakaan Lalu Lintas Jalan Raya di Kecamatan Pelaihari. (*JTIK*) *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5), 613-620. DOI: <http://dx.doi.org/10.25126/jtiik.2018551113> .
- [10] Wiza, F. (2019). Klasterisasi Karakteristik Kekerasan Seksual Terhadap Anak dengan Metode *K-Means Cluster Analysis*. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 10(1), 44-53. DOI: <https://doi.org/10.31849/digitalzone.v10i1.2423> .
- [11] Puntoriza, P., & Fibriani, C. (2020). Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means. *JOINS (Journal of Information System)*, 5(1), 86-94. DOI: <https://doi.org/10.33633/joins.v5i1.3469> .
- [12] Am, A., Defit, S., & Sumijan, S. (2019). Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering. *Jurnal KomtekInfo*, 5(3), 1-9.
- [13] Salam, A., Adiatma, D., & Zeniarja, J. (2020). Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS. *JOINS (Journal of Information System)*, 5(1), 62-68. DOI: <https://doi.org/10.33633/joins.v5i1.3350> .
- [14] Sarasvananda, I. B. G., Wardoyo, R., & Sari, A. K. (2019). *The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization*. *IJCCS (Indonesian Journal of Computing and Cybernetics System)*, 13(4), 313-322. DOI: <https://doi.org/10.22146/ijccs.45093> .
- [15] Annur, H. (2019). Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering. *JTU (Jurnal Informatika Upgris)*, 5(1), 40-45. DOI: <https://doi.org/10.26877/jiu.v5i1.3091> .
- [16] Hariyanto, M., & Shita, R. T. (2018). Clustering Pada Data Mining untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak *Euclidean Distance*. *SKANIKA (Sistem Komputer dan Teknik Informatika)*, 1(1), 117-122.
- [17] Nepal, B., Yamaha, M., Sahashi, H., & Yokoe, A. (2019). Analysis of Building Electricity Use Pattern Using K-Means Clustering Algorithm by Determination of Better Initial Centroids and Number of Clusters. *Energies*, 12(12), 1-17. DOI: <https://doi.org/10.3390/en12122451> .
- [18] *K-Means Advantages and Disadvantages* [online] (diperbaharui 10 Februari 2020) <https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages> [Diakses 06 September 2020].