

**KLASTERISASI PASIEN BPJS DENGAN METODE K-MEANS
CLUSTERING GUNA MENUNJANG PROGRAM JAMINAN
KESEHATAN NASIONAL DI RUMAH SAKIT ANWAR MEDIKA
BALONG BENDO SIDOARJO**

***CLUSTERING BPJS'S INPATIENTS BY USING K-MEANS CLUSTERING
METHOD TO SUPPORT THE NATIONAL HEALTH INSURANCE
PROGRAM AT ANWAR MEDIKA HOSPITAL
BALONG BENDO SIDOARJO***

¹Amir Ali*, ¹Lilis Masyfufah

¹Stikes Yayasan Rumah Sakit Dr. Soetomo

Info Artikel

Sejarah Artikel:

Submitted: 28

Nov 2020

Accepted: 19 Feb
2021

Publish Online: 19
Feb 2021

Kata Kunci:

SIRS, Database,
Data Mining,
K-Means, BPJS

Abstrak

Latar belakang: Manajemen RS Anwar Medika mengalami kesulitan dengan tumpukan data di dalam database sehingga database kurang optimal dalam penggunaannya. **Tujuan:** Menemukan informasi baru dari data rekam medis rawat inap pasien BPJS. **Metode:** Penelitian ini merupakan deskriptif kuantitatif. Data didapatkan dengan melakukan pengelompokan data dengan metode teknik data mining dengan menggunakan algoritma K-Means. **Hasil:** Menghasilkan 3 cluster; cluster 1 terdiri dari 91 pasien perempuan (51%), cluster 2 terdiri dari 26 pasien perempuan (14%), dan cluster 3 terdiri dari 63 pasien laki-laki (35%). Kecamatan asal pasien yang terbanyak berasal dari Krian dan Balongbendo. Diagnosis yang banyak diderita adalah Born in Hospital, Diarrhoea, dan Hemorrhagic. **Kesimpulan:** Banyak pasien BPJS didiagnosis dengan kode ICD X Z38.0, A09+E86, dan A91, sehingga bisa digunakan untuk penentuan anggaran rumah sakit. **Saran:** Penelitian selanjutnya dapat menggunakan kolaborasi dengan metode clustering hierarki agar didapatkan hasil pengelompokan data yang lebih baik.

Abstract

Background: The management of the RS Anwar Medika was experiencing difficulties with piles of data in the database so that the database is not optimally used. **Purpose:** Find new information from the medical record data of BPJS patients. **Method:** This research is descriptive quantitative. Data compile by grouping data using data mining techniques by using the K-Means algorithm. **Results:** 3 clusters where in cluster 1 consisted of 91 female patients (51%), cluster 2 consisted of 26 female patients (14%), and cluster 3 consisted of 63 male patients (35%). The sub-districts of origin the most patients came from the Krian and Balongbendo. The disease diagnosis, the patients suffered from a lot of born in hospital, Diarrhoea and Hemorrhagic diseases. **Conclusion:** Many BPJS patients were diagnosed with the disease with ICD X was Z38.0, A09+E86, and A91, so it can to arrange budget claims of hospital. **Suggestion:** For further research, it can use collaboration with the hierarchical clustering method in order to get better data grouping results

Keywords :

SIRS, Database,
Data Mining,
K-Means, BPJS,
universal coverage

PENDAHULUAN

Dalam era pembiayaan kesehatan merupakan bagian yang penting dalam implementasi UU Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional (SJSN), yang selanjutnya akan diselenggarakan oleh Undang-Undang Republik Indonesia (UURI) Nomor 24 Tahun 2011 tentang Badan Penyelenggara Jaminan Sosial (BPJS). BPJS adalah badan hukum yang dibentuk untuk menyelenggarakan program jaminan sosial. Dengan adanya program BPJS Kesehatan maka seluruh lapisan masyarakat memperoleh hak layanan kesehatan dari rumah sakit maupun puskesmas yang telah ditunjuk oleh pemerintah sebagai rumah sakit ataupun puskesmas pemberi layanan BPJS Kesehatan (Parasian, 2016). Pemerintah pusat maupun daerah saat ini sedang meningkatkan upaya penjaminan kesehatan bagi seluruh lapisan masyarakat. Upaya peningkatan sistem jaminan kesehatan berupa perubahan mendasar seperti penataan standarisasi pelayanan, tarif standar yang terjangkau, dan mudahnya akses kesehatan (Puput, 2017)

Perkembangan teknologi informasi sekarang ini melaju sangat pesat dan cepat menyebar ke berbagai bidang salah satunya di bidang kesehatan. Perkembangan teknologi informasi sangat dirasakan pada bidang kesehatan khususnya di suatu rumah sakit (Jerhi, 2015). Sejalan dengan Permenkes RI Nomor 1171/MENKES/PER/VI/2011 tentang Sistem Informasi Rumah Sakit (SIRS) dinyatakan bahwa setiap rumah sakit wajib melaksanakan SIRS sebagai proses pengumpulan, pengolahan, dan penyajian data rumah sakit. Di dalam SIRS terdapat database pasien. Database pasien ini pada dasarnya akan melakukan pencatatan yang dimulai dari pendaftaran pasien, proses pemeriksaan sampai pasien tersebut pulang, baik pulang sembuh atau mati berlaku pasien rawat jalan ataupun rawat inap.

Dari uraian tersebut di atas manajemen mengalami kesulitan dengan kumpulan data di dalam database Selanjutnya semua data terkumpul di satu database, sehingga database kurang optimal dalam penggunaannya terutama terkait database yang jumlahnya cukup besar atau yang kita kenal dengan istilah *big data* di Rumah Sakit Anwar Medika Balongbendo Sidoarjo.

Jumlah data pasien yang tersimpan dalam database SIRS di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo sampai bulan maret tahun 2019 mencapai 64812 pasien, dimana data ini masih tercampur dengan pasien umum sehingga sulit untuk mendapatkan data yang sesuai permintaan BPJS dengan beberapa variable yang diinginkan yaitu nomor pasien, nama pasien, tanggal lahir, jenis kelamin, diagnosa dan tindakan pasien rawat inap, sedangkan database dalam SIRS berisi semua data pasien seluruhnya yang tidak digunakan dalam klaim pembayaran pasien BPJS.

Tujuan dari penelitian ini adalah untuk menemukan informasi baru dari data rekam medis pasien yang tersimpan dalam database SIRS dimana nantinya informasi yang dihasilkan dapat digunakan untuk pengambilan keputusan oleh pihak manajemen rumah sakit. Seperti yang kita ketahui bahwa rekam medis pasien adalah input dari database SIRS. Ragam bentuk catatan tindakan medis yang dilakukan dalam pelayanan kesehatan itulah yang disebut dengan rekam medis (Susilowati, 2018)

Urgensi dari penelitian ini adalah bagaimana menemukan informasi baru dari data rekam medis pasien yang tersimpan dalam database SIRS berdasarkan data rekam medis pasien yang berobat di Rumah Sakit Anwar Medika Balongbendo Sidoarjo.

Pada penelitian ini kami mencoba menemukan karakteristik penting dari data rekam medis pasien yang tersimpan dalam database untuk dianalisa dan diolah agar dapat ditemukan

pengetahuan baru yang manfaatnya dapat dirasakan oleh masyarakat yang berobat ke Rumah Sakit Umum Anwar Medika Balongbendo Sidoarjo. Pihak rumah sakit dapat juga mengambil manfaatnya dengan memberikan penyuluhan terhadap daerah yang banyak terjangkit penyakit.

METODE PENELITIAN

Penelitian ini dilaksanakan di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. Pendekatan yang dilakukan adalah pendekatan dengan mengelompokkan data rekam medis pasien terutama pasien BPJS rawat inap dengan teknik data mining menggunakan metode K-means clustering sebagai teknik analisis datanya. Data *mining* merupakan sebuah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database yang besar (Daniel, 2005). Berikut desain penelitian yang dilakukan :



Gambar 1. Desain Penelitian

Data rekam medis yang digunakan yaitu norm, nama pasien, jenis kelamin, umur pasien, asal kecamatan pasien, diagnosa penyakit pasien serta tindakan yang diberikan. Data yang digunakan adalah data rekam medis pasien BPJS rawat inap triwulan pertama pada tahun 2019 yang berobat di Rumah Sakit Anwar Medika Balongbendo Sidoarjo dimana setelah dilakukan *cleaning* data diperoleh datanya sebanyak 180 data rekam medis pasien BPJS rawat inap.

Pada tahapan penelitian diatas, teknik pengumpulan dan instrumen penelitian diambilkan dari data rekam medis pasien berupa file excel yang di export dari database SIMRS. Kemudian dilakukan pengelompokan data/ *clustering* data dengan menggunakan teknik data mining dengan algoritma *K-Means*. *K-Means Clustering* merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007). Algoritma Kmeans terdapat aturan dimana :

- Berapa jumlah cluster yang perlu dimasukkan
- Hanya memiliki atribut bertipe numerik.

Algoritma *K-Means* pada dasarnya melakukan 2 proses yakni proses pendeteksian lokasi pusat cluster dan proses pencarian anggota dari tiap-tiap cluster. Proses clustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1,...,n$; $j=1,...,m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap kluster ditetapkan secara bebas (sembarang), C_{kj} ($k=1,...,k$; $j=1,...,m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k), dapat digunakan formula Euclidean. Suatu data akan menjadi anggota dari cluster ke- k apabila jarak data tersebut ke pusat cluster ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lain (Agustina, 2012)

Dalam mengukur jarak antara titik pusat atau *centroid* dengan data adalah dengan menggunakan *euclidean distance* (Fayyad, 1996). Dimana Rumus *euclidean distance* adalah

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2}$$

Dengan:

d_k = jarak antara data ke- i dengan titik pusat klaster ke- k

m = jumlah atribut

x_j = data ke- i

c_k = data pusat klaster ke- k

HASIL PENELITIAN

Berikut visualisasi proses clustering dengan menggunakan tools weka untuk atribut jenis kelamin, kecamatan asal pasien, dan diagnosa penyakit



Gambar 2. Visualisasi Clustering Data Pasien BPJS Rawat Inap untuk Atribut Jenis Kelamin



Gambar 1. Visualisasi *Clustering* Data Pasien BPJS Rawat Inap untuk Atribut Kecamatan Asal Pasien



Gambar 4. Visualisasi *Clustering* Data Pasien BPJS Rawat Inap untuk Atribut Diagnosa Penyakit

Dari informasi diatas, maka didapatkan hasil *Clusterisasi* datanya seperti pada tabel dibawah ini:

Tabel 1. Hasil Clustering Data

Hasil Cluster 1	Hasil Cluster 2	Hasil Cluster 3
Terdiri dari 91 orang yang berasal dari :	Terdiri dari 26 orang yang berasal dari :	Terdiri dari 63 orang yang berasal dari :
Kecamatan Asal Pasien :	Kecamatan Asal Pasien :	Kecamatan Asal Pasien :
1. Krian = 36	1. Krian = 7	1. Krian = 21
2. BalongBendo = 20	2. BalongBendo = 7	2. BalongBendo = 17
3. Tarik = 12	3. Tarik = 2	3. Tarik = 8
4. Taman = 9	4. Taman = 1	4. Taman = 6
5. Prambon = 8	5. Prambon = 4	5. Prambon = 3
6. Wonoayu = 1	6. Wonoayu = 3	6. Wonoayu = 4
7. Sukodono = 2	7. Sukodono = 2	7. Sukodono = 1
8. Tulangan = 2		8. Tulangan = 3
9. Sidoarjo = 1		
Diagnosa Penyakit menurut ICD terminologi :	Diagnosa Penyakit menurut ICD terminologi :	Diagnosa Penyakit menurut ICD terminologi :
1. Born in hospital = 14	1. Abdomen pain = 1	1. Born in hospital = 1
2. Diabetes Mellitus = 11	2. Chronic bronchitis = 1	2. Diabetes Mellitus = 4
3. Hemorrhagic = 5	3. Chronic renal failure = 1	3. Hemorrhagic = 10
4. Intestinal infection disease = 7	4. connective and other soft issue = 1	4. Intestinal infection disease = 5
5. Delivery by sc = 11	5. Cysta bartholin = 1	5. Single spontaneous delivery = 1
6. Single spontaneous delivery = 9	6. Dengue fever = 1	6. Acute bronchitis = 3
7. Acute bronchitis = 4	7. Diarrgoea = 2	7. Pharyngitis = 1
8. Pharyngitis = 5	8. Diarrhoea and gastroenteritis = 1	8. Concussion = 4
9. Concussion = 1	9. Disorder of pinna = 1	9. Diarrhoea = 1
10. Diarrhoea = 1	10. False labour = 1	10. Fracture = 4
11. Fracture = 1	11. Ganglion = 1	11. Dyspepsia = 3
12. Abortion incomplete = 4	12. Hipertensi = 1	12. Septicaemia = 2
13. Dyspepsia = 1	13. Liver = 1	13. Hearth disease = 2
14. Medical abortion = 4	14. Mononeuropathy = 1	14. Phymosis = 1
15. Septicaemia = 1	15. Myocardial infection = 1	15. Cataract = 1
16. Benign neoplasm = 2	16. Non Union Radius = 1	16. Cerbral infraction = 1
17. Blighted ovum = 2	17. Nontoxic goitre = 1	17. Chest pain = 1
18. Colic renal = 2	18. Open wound of ankle and foot = 1	18. Concussion + Fracture = 1
19. Exceccive mens + anemia = 2	19. Post partum = 1	19. Deglowing pedis = 1
20. Paratyphoid = 2	20. Premature reapture = 1	20. Diarrgoea = 2
21. Phymosis = 1	21. Pulmonary hearth disease = 1	21. Dizziness and giddiness (vertigo) = 1
22. Vomite earlier gravida = 1	22. renal colic + dengue haemorrhagic fever = 1	22. Fever = 1
Jenis Kelamin :	23. Spontaneous abortion = 1	23. Heaptic failure = 1
1. Perempuan = 91		24. Hepatomae = 1
2. Laki-Laki = 0		25. HIV = 1
		26. Lipoma = 1
		27. Mediastinum = 1
		28. Nefrolithiasis = 1
		29. Stroke infark = 1
		30. TBC Lung = 1
		31. Typhoid = 1

Jenis Kelamin : 1. Perempuan = 26 2. Laki-Laki = 0	32. Urinary tract infection = 1
	33. ventricular premature depolarization = 1
	34. Vomiting = 1
	Jenis Kelamin : 1. Perempuan = 0 2. Laki-Laki = 63

Dari tabel diatas didapatkan informasi bahwa pada hasil cluster 1 untuk kecamatan asal pasien banyak berasal dari kecamatan krian sebanyak 36 pasien. Untuk diagnosa penyakit menurut icd terminologi yang terbanyak adalah born in hospital sebanyak 14 pasien. Untuk jenis kelamin banyak didominasi oleh pasien dengan jenis kelamin perempuan sebanyak 91 pasien.

Sedangkan pada hasil cluster 2 didapatkan informasi untuk kecamatan asal pasien banyak berasal dari kecamatan krian dan balongbendo, masing-masing sebanyak 7 pasien. Untuk diagnosa penyakit menurut icd terminologi yang terbanyak adalah Diarrhoea sebanyak 2 pasien. Untuk jenis kelamin banyak didominasi oleh pasien dengan jenis kelamin perempuan sebanyak 26 pasien.

Dan pada hasil cluster 3 didapatkan informasi untuk kecamatan asal pasien banyak berasal dari kecamatan krian sebanyak 21 pasien. Untuk diagnosa penyakit menurut icd terminologi yang terbanyak adalah Hemorrhagic sebanyak 10 pasien. Untuk jenis kelamin banyak didominasi oleh pasien dengan jenis kelamin laki-laki sebanyak 63 pasien.

Data dari tabel hasil clustering diatas, maka didapatkan informasi juga bahwasannya kode ICD X dari diagnosa penyakit yang paling banyak baik pada cluster 1, cluster 2 dan cluster 3 adalah born in hospital dengan kode ICD X Z38.0, Diarrhoea dengan kode ICD X A09+E86, dan Hemorrhagic dengan kode ICD X A91. Dari data tabel di atas didapatkan juga informasi dimana terdapat banyak pasien dengan jenis kelamin perempuan yang menderita sakit dibandingkan dengan pasien dengan jenis kelamin laki-laki.

Bila informasi diatas dikaitkan dengan pembayaran klaim BPJS ke pemerintah maka, besaran tarif nya sebagai berikut:

Tabel 2. Biaya/ Tarif Klaim BPJS

Kode ICD X	Diagnosa Penyakit	Kelas	Tarif/ Biaya Klaim
ICD X Z38.0	born in hospital	a. Z38.0 BB kurang dari sama dengan 2499 gram maka untuk :	
		Kelas 1	
		Kelas 2	Rp 7,195,000
		Kelas 3	Rp 6,167,100
			Rp 5,139,300
		b. Z38.0 BB lebih dari sama dengan 2500 gram maka untuk :	
		Kelas 1	
		Kelas 2	Rp 4,796,700
		Kelas 3	Rp 4,111,400
A09+E86	Diarrhoea		Rp 3,426,200
		Kelas 1	Rp 2,027,800
		Kelas 2	Rp 1,738,100
		Kelas 3	Rp 1,448,400
A91	Hemorrhagic	Kelas 1	Rp 2,176,200
		Kelas 2	Rp 1,865,300
		Kelas 3	Rp 1,554,400

Sumber : Sistem CaseMix RS Anwar Medika Balongbendo Sidoarjo 2019

Dari informasi tabel tarif /biaya pengajuan klaim BPJS diatas dimana untuk pasien BPJS rawat inap di Rumah Sakit Anwar Medika Balongbendo Sidoarjo rata-rata mengambil biaya tarif/biaya klaim untuk kelas 3 jadi untuk diagnosa penyakit born in hospital dengan ICD X Z38.0 kelas 3 pembiayaan klaim BPJS nya sebesar Rp 5,139,300 untuk BB kurang dari sama dengan 2499 gram dan sebesar Rp 3,426,200 untuk BB lebih dari sama dengan 2500 gram. Untuk diagnosa penyakit Diarrhoea dengan kode ICD X A09+E86 Kelas 3 pembiayaan klaim BPJS nya sebesar Rp 1,448,400. Sedangkan diagnosa penyakit Hemorrhagic dengan kode ICD X A91 kelas 3 pembiayaan klaim BPJS nya sebesar Rp 1,554,400.

PEMBAHASAN

Pada proses clustering ini, akan dilakukan proses pengolahan data terlebih dahulu, dari data yang didapatkan dari rumah sakit setelah dilakukan pembersihan data yaitu dengan melihat kelengkapan data pasien BPJS ternyata didapatkan data pasien BPJS pada kurun waktu triwulan pertama di tahun 2019 yaitu sebanyak 180 data pasien BPJS.

Tabel 3. Data Pasien BPJS Triwulan Pertama Tahun 2019

No	NoRM	Nama Pasien	Jenis Kelamin	Asal Kecamatan Pasien	ICD menurut Terminologi
1.	5950XX	NAN	L	BALONGBENDO	Urinary tract infection
2.	5950 XX	RAN	L	PRAMBON	Hemorrhagic
3.	5949 XX	HIN	P	KRIAN	Liver
4.	5830 XX	NAB	P	KRIAN	Vomit earlier gravida
5.	5951 XX	SUL	P	WONOAYU	Hipertensi
6.	0784 XX	JOI	L	BALONGBENDO	HIV
7.	5125 XX	POE	L	WONOAYU	Diarrhoe
.....
180	6037 XX	MOK	L	WONOAYU	Concussion

Peneliti menggunakan 5 variabel pada data pasien BPJS rawat inap yang diperoleh yaitu NoRM, Nama Pasien, jenis kelamin, Asal Kecamatan Pasien dan ICD menurut terminologi. Sebelum diproses dengan clustering data menggunakan K-Means, maka data perlu dilakukan transformasi data terlebih dahulu.

A. Transformasi Data

Tabel tranformasi data diatas diperoleh dengan langkah-langkah sebagai berikut :

1. Urutkan data berdasarkan frekuensi kemunculannya
2. Inisialisasi data tersebut mulai dari data tertinggi dengan nilai 1, kemudian data selanjutnya 2,3 dan seterusnya

Berdasarkan data pada tabel 3, maka inisialisasi data untuk variabel jenis kelamin, asal kecamatan pasien, dan ICD menurut terminologi adalah

Tabel 4. Inisial Variabel Jenis Kelamin

Jenis Kelamin	Frekuensi	Inisial
Perempuan	117	1
Laki-laki	63	2

Tabel 5. Inisial Variabel Asal Kecamatan Pasien

Kecamatan Asal Pasien	Frekuensi	Inisial
Krian	64	1
Balongbendo	43	2
Tarik	22	3
Taman	16	4
Prambon	15	5
Wonoayu	8	6
Sukodono	5	7
Tulangan	4	8
Buduran	1	9
Sidoarjo	1	10
Waru	1	11

Tabel 6. Inisial Variabel ICD menurut Terminologi

Diagnosa Penyakit By ICD Terminologi	Frekuensi	Inisial
Born in hospital	15	1
Diabetes Mellitus	15	2
Hemorrhagic	15	3
Intestinal infection disease	12	4
Delivery by sc	11	5
Single spontaneous delivery	10	6
Acute bronchitis	7	7
Pharyngitis	6	8
Concussion	5	9
Diarrhoea	5	10
Fracture	5	11
Abortion incomplete	4	12
Dyspepsia	4	13
Medical abortion	4	14
Septicaemia	3	15
Benign neoplasm	2	16
Blighted ovum	2	17
Colic renal	2	18
Excessive menses + anemia	2	19
Heart disease	2	20
Paratyphoid	2	21
Phymosis	2	22
Vomiting earlier gravida	2	23
Abdomen pain	1	24
Cataract	1	25
Cerebral infarction	1	26
Chest pain	1	27
Chronic bronchitis	1	28

Diagnosa Penyakit By ICD Terminologi	Frekuensi	Inisial
Chronic renal failure	1	29
Concussion + Fracture	1	30
connective and other soft issue	1	31
Cysta bartholin	1	32
Deglowing pedis	1	33
Dengue fever	1	34
Diarrhoea	1	35
Diarrhoea and gastroenteritis	1	36
Disorder of pinna	1	37
Dizziness and giddiness (vertigo)	1	38
False labour	1	39
Fever	1	40
Ganglion	1	41
Heaptic failure	1	42
Hepatoma	1	43
Hipertensi	1	44
HIV	1	45
Lipoma	1	46
Liver	1	47
Mediastinum	1	48
Mononeuropathy	1	49
Myocardial infection	1	50
Nefrolithiasis	1	51
Non Union Radius	1	52
Nontoxic goitre	1	53
Open wound of ankle and foot	1	54
Post partum	1	55
Premature reapture	1	56
Pulmonary hearth disease	1	57
renal colic + dengue haemorrhagic fever	1	58
Spontaneous abortion	1	59
Stroke infark	1	60
TBC Lung	1	61
Trachea, bronchus and lung	1	62
Typhoid	1	63
Urinary tract infection	1	64
ventricular premature depolarization	1	65
Vomiting	1	66

Sehingga data hasil transformasi dapat dilihat pada tabel 7. berikut ini :

Tabel 7. Data Pasien BPJS Triwulan Pertama Tahun 2019 Hasil Transformasi

No	NoRM	Nama Pasien	Jenis Kelamin	Kecamatan Asal Pasien	ICD menurut Terminologi
1.	5950XX	NAN	2	2	64
2.	5950XX	RAN	2	5	3
3.	5949XX	HIN	1	1	47
4.	5830XX	NAB	1	1	65
5.	5951XX	SUL	1	6	44
6.	0784XX	JOI	2	2	45
7.	5125XX	POE	2	6	35
.....
180	6037XX	MOK	2	6	9

Hal ini seperti dapat kita lihat pada salah satu langkah pada penelitian terdahulu dengan judul penelitian “Klasterisasi Data Pasien Asuransi Kesehatan Menggunakan Metode K-Means (Studi kasus : RSUP H. Adam Malik Medan)” dari Rudolfo Damanik dan Rocky SN Nainggolan dimana peneliti tersebut menggunakan variabel yang dibutuhkan sebagai dasar melakukan proses *clustering* yaitu umur, jenis kelamin, kodya dan diagnosa penyakit. Dimana untuk melakukan Untuk transformasi data dapat dilakukan dengan langkah-langkah sebagai berikut:

1. Urutkan data berdasarkan frekuensi kemunculannya
2. Inisialisasikan data tersebut mulai dari data tertinggi dengan nilai 1, kemudian data selanjutnya 2, 3 dan seterusnya (Rudolfo, 2017)

B. Clustering Data

Setelah dilakukan proses transformasi data maka dilakukan proses clustering data yaitu dengan proses clusterisasi data ini, dimulai dari menentukan jumlah cluster yang akan dibentuk. Dari ke 180 data pasien BPJS tersebut, maka akan dikelompokkan ke dalam 3 cluster

Untuk penentuan mengelompokkan data pasien BPJS dengan 3 cluster dilakukan dengan menentukan initial cluster centre awal yang dipilih secara random. Lalu menghitung jarak data ke pusat cluster centre. Persamaan yang digunakan untuk menghitung jarak pada penelitian ini adalah Euclidean Distance yang memiliki persamaan sebagai berikut :

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$$

Pada proses clustering ini, untuk initial cluster centre awal yang dipilih secara random, didapatkan datanya sebagai berikut :

Untuk :

1. Cluster 1 : 1,1,2
2. Cluster 2 : 1,2,2
3. Cluster 3 : 2,3,9

Contoh perhitungan jarak data ke-1 ke pusat cluster pertama, cluster kedua dan cluster ketiga pada iterasi ke-1 adalah :

$$\begin{aligned} D(x_1, c_1) &= \sqrt{(\text{Inisiasi Jenis Kelamin} - \text{Pusat cluster1 untuk inisiasi jenis kelamin})^2 + \\ &\quad (\text{Inisiasi Asal Kecamatan Pasien} - \text{Pusat cluster1 untuk inisiasi asal kecamatan pasien})^2 + (\text{Inisiasi ICD menurut terminologi} - \text{Pusat cluster1 untuk inisiasi ICD menurut terminologi})^2} \\ &= \sqrt{(2 - 1)^2 + (2 - 1)^2 + (64 - 2)^2} \\ &= 62,02 \end{aligned}$$

$$\begin{aligned} D(x_1, c_2) &= \sqrt{(\text{Inisiasi Jenis Kelamin} - \text{Pusat cluster2 untuk inisiasi jenis kelamin})^2 + \\ &\quad (\text{Inisiasi Asal Kecamatan Pasien} - \text{Pusat cluster2 untuk inisiasi asal kecamatan pasien})^2 + (\text{Inisiasi ICD menurut terminologi} - \text{Pusat cluster2 untuk inisiasi ICD menurut terminologi})^2} \\ &= \sqrt{(2 - 1)^2 + (2 - 2)^2 + (64 - 2)^2} \\ &= 62,01 \end{aligned}$$

$$\begin{aligned} D(x_1, c_3) &= \sqrt{(\text{Inisiasi Jenis Kelamin} - \text{Pusat cluster 3 untuk inisiasi jenis kelamin})^2 + \\ &\quad (\text{Inisiasi Asal Kecamatan Pasien} - \text{Pusat cluster 3 untuk inisiasi asal kecamatan pasien})^2 + (\text{Inisiasi ICD menurut terminologi} - \text{Pusat cluster 3 untuk inisiasi ICD menurut terminologi})^2} \\ &= \sqrt{(2 - 2)^2 + (2 - 3)^2 + (64 - 9)^2} \\ &= 55,01 \end{aligned}$$

Cara yang sama dilakukan untuk data ke-2 sampai data ke-180, untuk dihitung jaraknya ke pusat cluster pertama, cluster kedua dan cluster ketiga
Sehingga didapatkan jaraknya sebagai berikut :

Tabel 8. Data Perhitungan Jarak dari data ke1 ke Pusat Masing-Masing Cluster

NoRM	Nama	Jenis Kelamin	Kecamatan Asal Pasien	ICD menurut Terminologi	Jarak ke		
					C1	C2	C3
5950XX	NAN	2	2	64	62.02	62.01	55.01
5950 XX	RAN	2	5	3	4.24	3.32	6.32
5949 XX	HIN	1	1	47	45.00	45.01	38.07
5830 XX	NAB	1	1	65	63.00	63.01	56.04
5951 XX	SUL	1	6	44	42.30	42.19	35.14
0784 XX	JOI	2	2	45	43.02	43.01	36.01
5125 XX	POE	2	6	35	33.39	33.26	26.17
6037 XX	MOK	2	6	9	8.66	8.12	3.00

Setelah itu baru data dikelompokkan menurut jarak terpendek ke pusat cluster baik cluster pertama, kedua dan ketiga seperti tabel di bawah ini :

Tabel 9. Data hasil pengelompokkan data ke cluster yang terdekat

NoRM	Nama	Jenis Kelamin	Kecamatan Asal Pasien	ICD menurut Terminologi	Jarak ke		
					C1	C2	C3
5950XX	NAN	2	2	64			*
5950 XX	RAN	2	5	3		*	
5949 XX	HIN	1	1	47			*
5830 XX	NAB	1	1	65			*
5951 XX	SUL	1	6	44			*
0784 XX	JOI	2	2	45			*
5125 XX	POE	2	6	35			*
.....
6037 XX	MOK	2	6	9			*

Pada iterasi berikutnya, setelah mendapatkan datanya. Maka dilakukan proses perhitungan kembali titik pusat cluster yang baru dengan rumus sebagai berikut :

$$C_i = \frac{1}{M} \sum_{j=1}^M X_j$$

Dimana : dilakukan penjumlahan data yang menjadi anggota clusternya dibagi sebanyak data anggota clusternya, maka akan didapatkan titik pusat cluster yang baru

Misalnya dari pengelompokan ke 180 data diatas, maka titik pusat cluster yang baru adalah :

Untuk Cluster 1 = (1.34, 2.63, 16.20)

Data diatas didapatkan dari penjumlahan data yang menjadi anggota clusternya dibagi sebanyak data anggota clusternya. Datanya sebagai berikut :

Tabel 10. Data Hasil Clustering Pada Iterasi -1 Untuk Cluster 1

NoRM	Nama	Jenis Kelamin	Kecamatan Asal Pasien	ICD menurut Terminologi
5929XX	MOC	2	1	2
5951XX	IIS	1	1	1
5399 XX	MIK	1	1	4
5967 XX	SUB	2	1	2
5658 XX	UMI	1	1	2
5417 XX	FIK	1	1	5
5973 XX	SYA	1	1	3
.....
6031 XX	DIC	2	1	3

Untuk Cluster 2 = (1.34, 2.71, 16.81)

Data diatas didapatkan dari penjumlahan data yang menjadi anggota clusternya dibagi sebanyak data anggota clusternya. Datanya sebagai berikut :

Tabel 11. Data Hasil Clustering Pada Iterasi -1 Untuk Cluster 2

NoRM	Nama	Jenis Kelamin	Kecamatan Asal Pasien	ICD menurut Terminologi
5950XX	RAN	2	5	3
5955 XX	GIL	2	5	4
5956 XX	AHM	2	7	4
5958 XX	MAS	1	4	5
5960 XX	SAL	1	2	2
5958 XX	AIN	1	5	2
5961 XX	SUT	2	8	2
5969 XX	ECI	1	3	5
.....
6037 XX	CHU	1	8	1

Untuk Cluster 3 = (1.35, 2.72, 17.03)

Data diatas didapatkan dari penjumlahan data yang menjadi anggota clusternya dibagi sebanyak data anggota clusternya. Datanya sebagai berikut:

Tabel 12. Data Hasil Clustering Pada Iterasi -1 Untuk Cluster 3

NoRM	Nama	Jenis Kelamin	Kecamatan Asal Pasien	ICD menurut Terminologi
5950XX	NAN	2	2	64
5949 XX	HIN	1	1	47
5830 XX	NAB	1	1	65
5951 XX	SUL	1	6	44
0784 XX	JOI	2	2	45
5125 XX	POE	2	6	35
5953 XX	FAN	2	3	30
.....
6037 XX	MOK	2	6	9

Tabel 13. Tabel Pusat Cluster Baru

Cluster 1	1.34	2.63	16.20
Cluster 2	1.34	2.71	16.81
Cluster 3	1.35	2.72	17.03

Iterasi selanjutnya dilakukan dengan cara yang sama seperti cara diatas hingga tidak ada perubahan data dalam suatu cluster (Wardhani,2016). Dari hasil penelitian didapatkan hasil yaitu pada iterasi ke-11 iterasi berhenti karena tidak ada perubahan data dalam clusternya. Hal sejalan dengan penelitian sebelumnya dari anindya khrisna wardhani dengan judul penelitiannya “Implementasi algoritma k-means untuk pengelompokan penyakit pasien pada puskesmas kaje pekalongan” dimana anggota data dalam pengelompokan data tidak berubah lagi setelah dilakukan perhitungan jarak data dengan pusat clusternya.

SIMPULAN

Hasil identifikasi data rekam medis pasien BPJS pada triwulan pertama tahun 2019 didapatkan sebanyak 180 data pasien, dimana dalam menganalisa data rekam medis pasien BPJS dengan menggunakan algoritma *K-Means Clustering* dengan 3 cluster yang dibentuk.

Berdasarkan hasil clustering dengan algoritma K-Means, maka didapatkan hasil bahwa pasien yang paling banyak menderita sakit dari kecamatan krian dan balonbendo, untuk pasien dengan jenis kelamin perempuan lebih mendominasi banyak penyakitnya daripada jenis kelamin laki-laki. Sedangkan diagnosa penyakit berdasarkan ICD menurut terminologi lebih banyak didiagnosa born in hospital dengan kode ICD X Z38.0, Diarrgoea dengan kode ICD X A09+E86, dan Hemorrhagic dengan kode ICD X A91. 5.

Rumah sakit mendapatkan informasi bahwasannya diagnosa born in hospital dengan kode ICD X Z38.0, Diarrgoea dengan kode ICD X A09+E86, dan Hemorrhagic dengan kode ICD X A91 merupakan kode penyakit yang banyak terjadi pada pasien. Hal ini jika dikaitkan dengan grouping pada kode INA-CBG, maka tarif dengan kode ICD X diataslah yang paling banyak diderita oleh pasien BPJS. Dari informasi tabel tarif /biaya pengajuan klaim BPJS dimana untuk pasien BPJS rawat inap di Rumah Sakit Anwar Medika Balongbendo Sidoarjo rata-rata mengambil biaya tarif/biaya klaim untuk kelas 3 jadi untuk diagnosa penyakit born in hospital dengan ICD X Z38.0 kelas 3 pembiayaan klaim BPJS nya sebesar Rp 5,139,300 untuk BB kurang dari sama dengan 2499 gram dan sebesar Rp 3,426,200 untuk BB lebih dari sama dengan 2500 gram. Untuk diagnosa penyakit Diarrgoea dengan kode ICD X A09+E86 Kelas 3 pembiayaan klaim BPJS nya sebesar Rp 1,448,400. Sedangkan diagnosa penyakit Hemorrhagic dengan kode ICD X A91 kelas 3 pembiayaan klaim BPJS nya sebesar Rp 1,554,400.

Pengelompokan data rekam medis pasien BPJS dari proses data mining diatas adalah untuk menghasilkan informasi baru mengenai pola pengelompokan data pasien BPJS di kabupaten sidoarjo. Hal ini dapat dijadikan acuan bagi rumah sakit untuk melakukan sosialisasi terkait pencegahan penyakit pada kecamatan asal pasien yang banyak menderita sakit

SARAN

Dari hasil penelitian didapatkan bahwa dari informasi tabel tarif /biaya pengajuan klaim BPJS dimana untuk pasien BPJS rawat inap di Rumah Sakit Anwar Medika Balongbendo Sidoarjo rata-rata mengambil biaya tarif/biaya klaim untuk kelas 3. Hal ini perlu ditingkatkan dalam pelayanannya terhadap pasien BPJS rawat inap untuk kelas 3, agar pasien merasa puas dalam pelayanannya.

Untuk penelitian selanjutnya, peneliti lain dapat mengembangkan metode clustering yang berbeda pada penerapan klasterisasi pasien bpjs guna menunjang program jaminan kesehatan nasional

REFERENSI

- Agusta, Y, 2007, K-means - Penerapan, Permasalahan dan Metode Terkait, *Jurnal Sistem dan Informatika* Vol. 3 (Februari 2007) : 47-60
- A. T. dan F. K. S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto. 2012, "Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means," [Online]. Available:<http://yudistira.lecture.ub.ac.id/files/2014/04/clustering-kualitas-beras-dengan-k-means.pdf>.
- A. K. Wardhani. 2016. "Implementasi Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Kajen Pekalongan," *Transformatika*, vol. 14, p. 35
- D. S. Susilowati, Indah. Wisnaningsih Surjoseputro, "Perlindungan hukum terhadap hak privasi dan data medis pasien di rumah sakit x surabaya," *Wiyata*, vol. 5 No.1, 2018.
- J. W. Fernanda, "Sistem informasi peminjaman dokumen rekam medis di rumah sakit x," *Wiyata*, vol. 2 No.1, 2015
- Larose, Daniel. 2005. *Discovery Knowledge in Data. Canada* :A Jhon Wiley & Sons, Inc Publication
- P. Noviana, "Analisis tingkat kepuasan pasien rawat jalan terhadap kualitas pelayanan instalasi farmasi di rsud pare menggunakan metode servqual," *J. Wiyata*, vol. Volume 04, p. 111, 2017
- R. D. Silitonga, Parasian D P. 2016. "mplementasi algoritma k-means clustering pada analisis penyebaran penyakit pasien pengguna badan penyelenggara jaminan sosial kesehatan (bpjs) (studi kasus : rumah sakit umum pusat haji adam malik medan)," *J. Tek. Inform. Unika St. Thomas*, vol. 01, p. 38
- R. R. S. N. Damanik, "Klasterisasi Data Pasien Asuransi Kesehatan Menggunakan Metode K-Means (Studi Kasus RSUP H. Adam Malik Medan)," *Pelita Inform. Budi Darma*, vol. XVI, p. 184, 2017
- S. U, Fayyad, G, Piatetsky-Shapiro, P, *Advances in Knowledge Discovery and Data Mining*. 1996