



Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan *K-Means Clustering*

Muhammad Bhakti Fajri¹, Susan Dian Purnamasari²

^{1,2}Information System Department, Bina Darma University, Palembang, Indonesia
Email: ¹fajribhakti6@gmail.com, ²susandian@binadarma.ac.id

Abstrak

Terdapat catatan kunjungan pasien pada setiap poli sebagai arsip bagi pihak Puskesmas Muara Enim. Pada catatan tersebut memuat Informasi usia, jenis kelamin, jenis penyakit serta Poli pemeriksaan. Data-data pasien tersebut mengalami penumpukan sehingga sulit menarik kesimpulan untuk memperbaiki kualitas serta layanan. Tujuan dari penelitian ini adalah untuk mencari pola baru yaitu kecenderungan penyakit dengan usia melalui data pasien di Puskesmas Muara Enim dari tahun 2020-2021. Dengan menggunakan teknik *data mining* serta memanfaatkan algoritma *K-Means clustering* untuk mengelompokan usia pasien berdasarkan kelompok usia yang berpotensi dalam penyebaran penyakit pada Puskesmas Muara Enim. Data tersebut akan dilakukan *cleaning*, *selection* dan akan ditentukan jarak usia pasien yang akan di bagi menjadi 4 kelompok yaitu Anak-anak, Remaja, Dewasa dan Lansia, yang selanjutnya akan dilakukan *clustering* data. Dibantu dengan aplikasi *rapidminer*, data yang sudah diolah akan dihitung secara manual yang menghasilkan sebanyak total 5015 pasien ditahun 2020 dan 5466 pasien ditahun 2021. Berdasarkan data tersebut dapat disimpulkan bahwa kelompok rentang usia tertinggi yang mengidap penyakit adalah Pasien Dewasa dan Anak – Anak kemudian pasien lansia dan terakhir pasien remaja pada setiap tahunnya. Pada tahun 2020 dan 2021 penyakit Asma, Diare, Penyakit tekanan darah tinggi, Tuberkulosis, Imunisasi dan Demam, batuk pilek merupakan penyakit dengan penderita tertinggi.

Kata Kunci: *K-Means Clustering*, Puskesmas, Usia, Penyakit

1. PENDAHULUAN

Perkembangan teknologi saat ini sudah sangat pesat, tidak hanya pada bidang informasi, industri, pendidikan, budidaya tetapi pada bidang kesehatan. Para ahli terus mengembangkan teknologi yang ada, sehingga dengan adanya teknologi tersebut para ahli merasa terbantu dalam menyelesaikan pekerjaannya. Dengan berkembangnya teknologi tersebut diharapkan dapat mempermudah seseorang memperoleh pengetahuan tentang kesehatannya [1].



Puskesmas merupakan salah satu instansi kesehatan yang berada ditingkat Kabupaten, termasuk Puskesmas Muara Enim yang berada di Kabupaten Muara Enim, Puskesmas Muara Enim merupakan suatu unit pelaksana fungsional yang berfungsi sebagai pusat pembangunan kesehatan, pusat pembinaan peran serta masyarakat dalam bidang kesehatan serta pusat pelayanan kesehatan tingkat pertama yang menyelenggarakan kegiatannya secara menyeluruh, terpadu yang berkesinambungan pada suatu masyarakat khususnya di daerah Kabupaten Muara Enim [2].

Dalam rangka menciptakan pelayanan Instansi Kesehatan yang baik diperlukan tata kerja yang tertib, rapi, dan teliti sehingga menghasilkan informasi yang cepat, akurat, dan tepat waktu sesuai dengan kebutuhan. Seperti instansi kesehatan pada umumnya, Puskesmas Muara Enim juga memberikan pelayanan kesehatan bagi masyarakat setiap harinya. Masyarakat yang ingin berobat di Puskesmas Muara Enim terlebih dahulu melakukan pendaftaran pada bagian registrasi dan selanjutnya masyarakat tersebut akan diarahkan kepada Poli yang dituju berdasarkan usia pasien. Seiring dengan bertambahnya jumlah pasien tersebut, maka bertambah pula data pasien setiap harinya, sehingga data yang banyak tersebut dapat dimanfaatkan secara optimal oleh Puskesmas Muara Enim untuk dijadikan arsip setiap tahunnya dan hasilnya dapat dijadikan bahan atau dasar penyuluhan kesehatan oleh Puskesmas Muara Enim [3].

Metode *K-Means Clustering* merupakan metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengelompokan data yang mempunyai kemiripan karakteristik antara data satu dengan data lain yang telah diperoleh data yang memiliki kesamaan bukan data yang sama tetapi memiliki karakteristik yang sama [4]. Dengan menerapkan metode *K-Means Clustering* dapat membantu pihak pemerintah khususnya Puskesmas Muara Enim agar dapat mengetahui pasien apa yang sering terkena penyakit berdasarkan usianya.

Berdasarkan penelitian sebelumnya yang berjudul "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Anwar Medika Bolong Bendi Sidoarjo" pada tahun 2019 Data pasien yang tersimpan pada database aplikasi SIMRS di rumah sakit pada umumnya hanya dimanfaatkan untuk membuat laporan dan grafik pasien rumah sakit, data penyakit pasien serta biaya berobat pasien. Database yang ada belum dapat digunakan sebagai dasar dalam pengambilan keputusan. Dengan menggunakan metode k-means untuk membagi data menjadi subset data berdasarkan kesamaan atau kemiripan sehingga dalam menemukan informasi baru berdasarkan data rekam medis pasien yang tersimpan dalam database SIMRS (Sistem Informasi Manajemen Rumah Sakit) Anwar

Medika Sidoarjo agar dapat digunakan dalam proses pengambilan keputusan oleh pihak manajemen rumah sakit agar tepat sasaran. Peneliti dapat mengidentifikasi data rekam medis dari rumah sakit anwar medika sebanyak 534 data pasien dengan waktu penyelesaian sebanyak 0.06 detik oleh sistem. Pengelompokan data rekam medis pasien dari proses data mining diatas adalah untuk menghasilkan informasi baru mengenai pola pengelompokan penyebaran penyakit di setiap kecamatan [5].

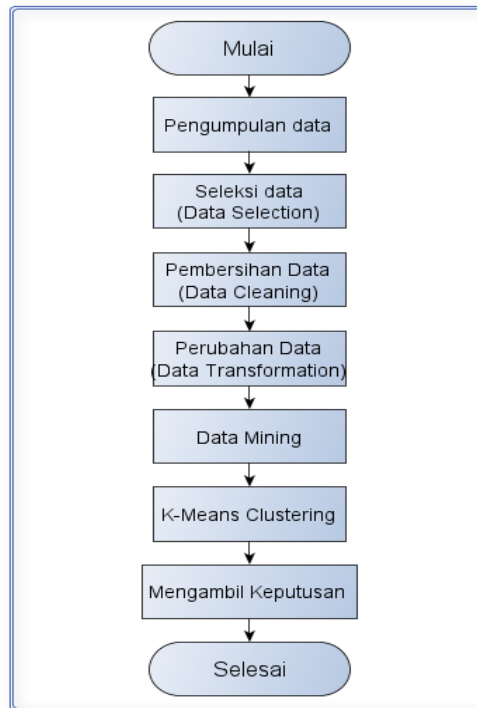
Terdapat pula penelitian dengan judul "Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering" yang menyatakan saat ini tumpukan data pasien yang ada di RSUD Pusat Haji Adam Malik Medan umumnya hanya dimanfaatkan pada pemenuhan laporan dan grafik pasien rumah sakit, data penyakit serta biaya perawatan pasien. Tumpukan data yang ada belum menyajikan pola khusus terhadap data seperti kecenderungan penyakit pasien dikaitkan dengan usia pasien. Dengan diketahuinya pola kecenderungan penyakit pasien dikaitkan dengan usia pasien, diharapkan pihak rumah sakit dapat melakukan program penyuluhan dengan tepat serta dapat melakukan tindakan antisipasi prioritas layanan. dari sejumlah pasien yang ada, persentasi usia pasien paling tinggi adalah pasien dengan usia tua dan kemudian pasien dengan usia parobaya. Melalui penelitian ini diharapkan dapat menjadi masukan bagi Dinas Kesehatan berkaitan dengan penyuluhan kesehatan bagi masyarakat [6].

Dari penelitian penelitian terdahulu dapat disimpulkan bahwa algoritma K-Means dapat mengelompokkan *dataset* yang belum mempunyai label sehingga dapat digunakan pada *dataset* kunjungan pasien Puskesmas Muara Enim dari tahun 2020-2021 untuk mencari pola kesaamaan antar data dengan menggunakan variabel usia, jenis kelamin dan jenis penyakit sehingga menghasilkan informasi baru yang dapat dimanfaatkan oleh pihak Puskesmas Muara Enim dalam pengambilan keputusan serta instansi terkait dalam menyelenggarakan penyuluhan khususnya didaerah Muara Enim.

2. METODOLOGI PENELITIAN

Algoritma K-Means merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat kluster (centroid) terdekat dengan data. Tujuan K-Means adalah mengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antara kluster. ukuran kemiripan yang digunakan dalam kluster adalah fungsi jarak. sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data titik centroid [7].

Ada beberapa tahapan yang akan dilakukan dalam penelitian klusterisasi pola penyebaran penyakit pasien berdasarkan usia menggunakan k-means clustering yang dapat dilihat sebagai berikut [8] :



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Penelitian ini mengambil data kunjungan pasien ditahun 2020-2021 dikarenakan penulis memerlukan data terkini dari objek penelitian yaitu Puskesmas Muara Enim. Data tersebut pula sudah dilakukan rekapitulasi serta digitalisasi oleh pihak Puskesmas Muara Enim, dimana pada data kunjungan pasien yang didapat dari Puskesmas Muara Enim memuat beberapa informasi seperti kode penyakit, jenis penyakit, rentang usia pasien, jenis kelamin, dan poli pemeriksaan, lalu data kunjungan pasien tersebut diurutkan berdasarkan bulan pada setiap *sheet* yang ada. Untuk proses analisa pada penelitian ini menggunakan teknik data mining yang memanfaatkan algoritma *K-Means clustering* dibantu dengan aplikasi *rapidminer*.

2.2 Data Selection

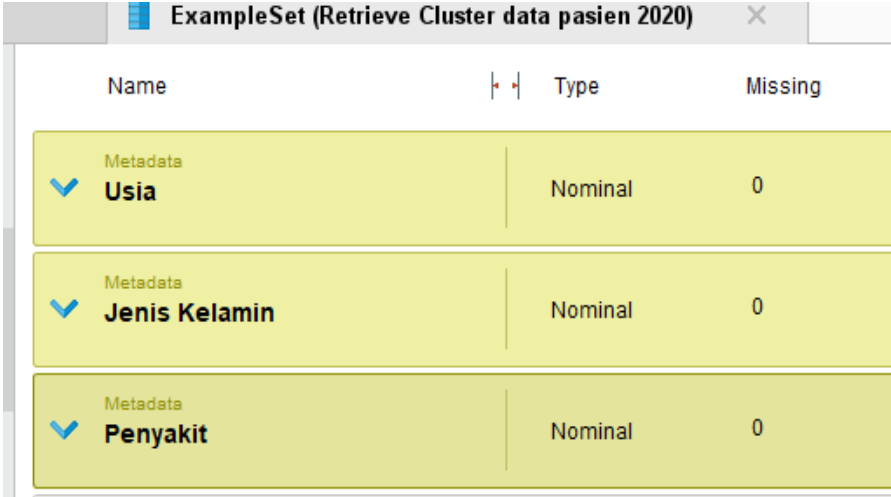
Dari data yang didapat dari objek penelitian berupa data kunjungan pasien di tahun 2020-2021 akan dilakukan seleksi data atau atribut yang relevan untuk penelitian. Sehingga didapatkan atribut yang akan diseleksi untuk penelitian ini yaitu atribut tahun, atribut bulan, atribut usia, atribut jenis kelamin dan atribut penyakit. Setelah menentukan seleksi atribut yang akan digunakan, maka data kunjungan pasien akan dikelompokkan berdasarkan kolom tahun, kolom bulan, kolom usia, kolom jenis kelamin dan kolom penyakit, dimana data kunjungan pasien yang dilakukan *data selection*, pada tahun 2020 terdapat 5015 data pasien dan tahun 2021 terdapat 5466 data pasien yang melakukan pemeriksaan pada Puskesmas Muara Enim sehingga didapatkan hasil *dataset* seperti pada gambar 2.

	A	B	C	D	E
1	Bulan	Tahun	Usia	Jenis Kelamin	Penyakit
2	Januari	2020	20-44 Tahun	L	Demam, batuk dan pilek
3	Januari	2020	20-44 Tahun	L	Demam, batuk dan pilek
4	Januari	2020	20-44 Tahun	L	Demam, batuk dan pilek
5	Januari	2020	20-44 Tahun	L	Demam, batuk dan pilek
6	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
7	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
8	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
9	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
10	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
11	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
12	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
13	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
14	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
15	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
16	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
17	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
18	Januari	2020	20-44 Tahun	P	Demam, batuk dan pilek
19	Januari	2020	20-44 Tahun	L	Demam, batuk dan pilek
20	Januari	2020	20-44 Tahun	L	Demam, batuk dan pilek

Gambar 2. *Dataset* kunjungan pasien

2.3 Data Cleaning

Kemudian setelah dilakukan *data selection*, pada tahapan ini dilakukan data *cleaning*, dimana nilai yang ada pada setiap atribut yang kosong akan dihapus agar meminimalisir terjadinya *error* pada saat proses *clustering* di aplikasi *rapidminer*, pada penelitian ini untuk *dataset* kunjungan pasien dari tahun 2020 dan 2021 tidak terdapat data dan atribut yang hilang atau *missing*. Dibuktikan dengan berhasilnya proses *input* pada aplikasi *rapidminer* yang menandakan bahwa tidak terdapat nilai yang hilang pada setiap atribut yang dapat dilihat pada gambar 3.



The screenshot shows a window titled "ExampleSet (Retrieve Cluster data pasien 2020)". It displays a table with three rows of metadata. Each row has a blue checkmark icon on the left, followed by the attribute name, its type, and the number of missing values.

Name	Type	Missing
Metadata Usia	Nominal	0
Metadata Jenis Kelamin	Nominal	0
Metadata Penyakit	Nominal	0

Gambar 3. *Cleaning Dataset* Kunjungan Pasien.

2.3 Data Transformation

Dalam proses *clustering* Algoritma *K-Means* hanya dapat menerima input data berupa angka. Oleh karena itu *Dataset* Kunjungan Pasien akan ditransformasi kedalam bentuk angka pada beberapa atribut yang terdapat huruf, dan dipilih 3 atribut yang digunakan dalam proses *clustering* yaitu atribut usia, jenis kelamin dan penyakit berikut adalah hasil *data transformation* pada penelitian ini :

a) Transformasi Rentang Usia

Departemen Kesehatan RI tahun 2009 membagi kelompok usia menjadi beberapak kategori diantaranya masa balita dari usia 0 – 5 tahun, masa kanak – kanak 5 – 11 tahun, masa remaja 12 -25 tahun, masa dewasa 25 – 45 tahun [9]. Dan menurut Peraturan Pemerintah Republik Indonesia Nomor 43 Tahun 2004 tentang pelaksanaan upaya peningkatan kesejahteraan sosial lanjut usia, Departemen Kesehatan RI tahun 2009 membagi kaelompok lansia menjadi 3 yaitu masa lansia awal 46- 55 tahun, masa lansia akhir 56 – 65 tahun, dan masa manula 65 – sampai atas [10]. Oleh karena itu transformasi rentang usia yang dapat dilakukan pada penelitian ini dengan berdasarkan dari kategori usia yang dikelompokkan oleh Departemen Kesehatan RI tahun 2009 dapat dilihat pada tabel 1.

Tabel 1 Transformasi data pada atribut usia

Rentang Usia	Kelompok Usia	Inisialisasi
0-7 hari		
8-28 Hari		
1 Bln - < 1 Tahun		
1-5 Tahun	Anak - Anak	1
6-9 Tahun		
10-14 Tahun		
15-19 Tahun	Remaja	2
20-44 Tahun	Dewasa	3
45-54 Tahun		
55-59 Tahun		
60-69 Tahun	Lansia	4
>70 Tahun		

Terdapat 12 kelompok rentang usia pasien yang ada pada Puskesmas Muara Enim yang kemudian dikelompokkan kedalam 4 kelompok yaitu anak – anak, remaja, dewasa dan lansia.

b) Transformasi Jenis Kelamin

Untuk atribut jenis kelamin yang merupakan data diskrit terdapat 2 entitas yaitu laki-laki dan perempuan. pada atribut jenis kelamin juga perlu merubah skala data kedalam bentuk angka sehingga data pada atribut jenis kelamin memiliki distribusi yang sama dengan nilai aslinya. Oleh karena itu ke-2 entitas tersebut dapat diwakilkan dengan angka yang dimana laki-laki diubah menjadi 1 sementara perempuan diubah menjadi 2 agar dapat diolah oleh algoritma K-Mean.

c) Transformasi Penyakit

Pada atribut penyakit dilakukan proses inisialisasi yang diurutkan berdasarkan banyaknya pasien yang menderita penyakit tersebut pada Puskesmas Muara Enim, dimana ada perbedaan jumlah penyakit yang diidap oleh pasien yang ada pada Puskesmas Muara Enim. Terdapat perbedaan Jenis penyakit antara tahun 2020 dan tahun 2021 dimana pada tahun 2020 terdapat 26 jenis penyakit dan pada tahun 2021 terdapat 25 jenis penyakit yang lebih jelasnya dapat di lihat pada Tabel 2.

Tabel 2 Transformasi data penyakit pasien tahun 2020 dan 2021

No	Penyakit	2020	2021
1	Keguguran	2	0
2	Pendarahan pada kehamilan, persalinan dan nifas	3	3
3	Kelainan Kornea dan refraksi	16	52
4	Gizi Buruk	21	10
5	Penyakit pulpa dan jaringan periapikal	22	26
6	Penyakit kulit infeksi	48	93
7	Katarak	49	27
8	Kesehatan Reproduksi	50	51
9	Demam Berdarah	71	62
10	Karies Gigi	75	100
11	Infeksi pada telinga	84	183
12	Penyakit rongga mulut, kelenjar ludah, rahang dan lainnya	94	85
13	Persalinan	97	133
14	PKB (Keluarga Berencana)	98	126
15	Penyakit mata lainnya	110	167
16	Cacar	113	115
17	Diabetes	119	83
18	Kolera	120	122
19	Disentri	129	148
20	Penyakit kulit Alergi	160	279
21	Asma	261	295
22	Diare	352	306
23	Penyakit tekanan darah tinggi	357	310
24	Tuberkolusis	532	531
25	Imunisasi	576	594
26	Demam, Batuk Pilek	1455	1566

Dapat dilihat pada tabel 2 bahwa beberapa penyakit yang diidap oleh pasien di Puskesmas Muara Enim mengalami peningkatan dari tahun 2020 hingga 2021. Beberapa contoh penyakit yang mengalami peningkatan secara signifikan adalah penyakit demam, batuk pilek, imunisasi, asma, penyakit kulit alergi, penyakit mata lainnya, persalinan, KB dan infeksi pada telinga. Serta diikuti oleh beberapa penyakit yang pula mengalami penurunan jumlah secara signifikan dari tahun 2020 hingga 2021 beberapa diantaranya adalah penyakit tekanan darah tinggi, diare, diabetes, katarak dan gizi buruk. Dari tabel 2 pula dapat dilihat bahwa untuk penyakit yang paling banyak diidap oleh pasien di Puskesmas Muara Enim baik di tahun 2020 dan 2021 yaitu demam, batuk, pilek disusul oleh imunisasi dan tuberkulosis. Untuk penyakit yang paling sedikit diidap oleh pasien di Puskesmas Muara Enim di tahun 2020 adalah keguguran, sementara di tahun 2021 penyakit

yang paling sedikit diidap oleh pasien di Puskesmas Muara Enim adalah Pendarahan pada kehamilan, persalinan dan nifas.

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Algoritma *K-Means* secara manual

Algoritma *K-Means* merupakan algoritma *unsupervised learning* yang dipakai untuk mengelompokkan *dataset* yang belum dilabel ke dalam *cluster* yang berbeda [11]. Dimana data yang ada akan dikelompokkan berdasarkan jumlah k atau *cluster* yang telah ditentukan. Dalam penerapan algoritma *K-Means* untuk *dataset* kunjungan pasien pada Puskesmas Muara Enim akan diambil sampel data dari *dataset* kunjungan pasien pada tahun 2020 secara tidak berurutan sebanyak 20 *record* dimana data tersebut telah di transformasi kedalam bentuk angka dan akan di kelompokkan kedalam 3 *cluster*.

Tabel 3 Sampel *dataset* kunjungan pasien tahun 2020

Pasien	Transformasi Usia	Transformasi Jenis Kelamin	Transformasi Penyakit
1	4	1	26
2	3	1	26
3	3	1	26
4	3	1	26
5	3	2	26
183	4	2	26
184	4	2	26
185	4	2	26
186	4	2	26
187	4	2	26
2736	3	2	1
2737	3	2	2
2851	4	2	3
3211	3	2	13
3212	3	2	13
3213	3	2	13
3214	4	2	13
3215	4	2	13
3332	4	1	3
4031	4	1	3

Secara umum penerapan algoritma *K-Means* dapat dilakukan dalam beberapa tahapan yang dapat dilihat sebagai berikut.

1. Tentukan jumlah *cluster* yang di bentuk. Dimana pada penelitian ini dipilih jumlah cluster sebanyak 3. yaitu *cluster* 1 merupakan *cluster* dengan penyakit

dan penderita tertinggi, *cluster 2* merupakan *cluster* dengan penyakit dan penderita sedang sementara pada *cluster 3* merupakan *cluster* dengan penyakit dan penderita sedikit.

2. Tentukan titik pusat *cluster* awal secara *random*. Dimana pada penelitian ini titik pusat *cluster* dipilih dari sampel data pasien ke 183 untuk *cluster 1* , sampel data pasien ke 3211 untuk *cluster 2* dan sampel data pasien ke 2736 untuk *cluster 3*.

Tabel 4 Titik pusat cluster

Titik pusat <i>cluster</i>	Transformasi usia	Transformasi Jenis Kelamin	Transformasi Penyakit
<i>Cluster 1</i>	4	2	26
<i>Cluster 2</i>	3	2	13
<i>Cluster 3</i>	3	2	1

3. Hitung jarak setiap objek pada setiap *centroid* dari masing-masing *cluster*. lalu hitung jarak antara objek dengan *centroid*, dengan menggunakan rumus *Euclidian Distance*.

$$d(q,r) = \sqrt{\sum_{i=1}^n (q_i - r_i)^2} \quad ; i=1,2,3,\dots,n$$

Dimana:

q : objek q ke-i

r : daya r ke-i

n : banyaknya objek

Berikut hasil perhitungan jarak dari sampel *dataset* kunjungan pasien tahun 2020 dengan titik pusat *cluster* pada iterasi pertama dengan rumus *Euclidian Distance*. Jarak data ke 1 pada tiap pusat *cluster* adalah:

$$\begin{aligned} d(1,1) &= \sqrt{(4-4)^2 + (1-2)^2 + (26-26)^2} \\ &= 1 \end{aligned}$$

Dan seterusnya, dilanjutkan menghitung data ke n pada tiap pusat cluster dan didapatkan seluruh hasil perhitungan pada Tabel 7.

Tabel 5 Hasil perhitungan sampel data pasien pada iterasi 1

Pasien	Jarak cluster 1	Jarak cluster 2	Jarak cluster 3	Hasil Cluster
1	1	13,0767	25,03997	1
2	1,414214	13,0384	25,01999	1
3	1,414214	13,0384	25,01999	1
4	1,414214	13,0384	25,01999	1
5	1	13	25	1
183	0	13,0384	25,01999	1
184	0	13,0384	25,01999	1
185	0	13,0384	25,01999	1
186	0	13,0384	25,01999	1
187	0	13,0384	25,01999	1
2736	25,01999	12	0	3
2737	24,02082	11	1	3
2851	23	10,04988	2,236068	3
3211	13,0384	0	12	2
3212	13,0384	0	12	2
3213	13,0384	0	12	2
3214	13	1	12,04159	2
3215	13	1	12,04159	2
3332	23,02173	10,0995	2,44949	3
4031	23,02173	10,0995	2,44949	3

Alokasikan masing-masing objek atau data ke dalam titik pusat *cluster* yang paling terdekat. Sehingga didapatkan sampel data dengan masing masing *cluster*, dimana data pasien 1, 2, 3, 4, 5, 183, 184, 185, 186, 187 merupakan anggota dari *cluster* 1 dan data pasien 2736, 2737, 2851, 3332, 4031 merupakan anggota dari *cluster* 3 kemudian data pasien 3211, 3212, 3213, 3214, 3215 merupakan anggota dari *cluster* 2.

- Setelah menentukan anggota tiap *cluster* kemudian dilakukan penentuan titik pusat *cluster* baru. Titik pusat *cluster* yang baru didapatkan dari perhitungan nilai rata rata dari setiap anggota *cluster* pada iterasi pertama dengan rumus sebagai berikut :

$$cb = \frac{\sum_{i=1}^n q_i}{n} ; i=1,2,3,\dots,n$$

Penentuan titik pusat *cluster* baru :

$$C1 = \frac{4+3+3+3+3+4+4+4+4+4}{10} = 3,6$$

$$= \frac{1+1+1+1+2+2+2+2+2+2}{10} = 1,6$$

$$= \frac{26+26+26+26+26+26+26+26+26+26}{10} = 26$$

Dan seterusnya sehingga didapatkan hasil perhitungan nilai titik pusat *cluster* yang baru pada Tabel 6.

Tabel 6 Titik pusat *cluster* baru

Cluster 1	3,6	1,6	26
Cluster 2	3,4	2	13
Cluster 3	3,4	1,6	9

- Ulangi kembali perhitungan jarak yang sama dengan menggunakan rumus *Euclidian Distance* dan titik pusat *cluster* yang baru sehingga data pasien pada setiap anggota *cluster* tidak berubah lagi. Pada sampel *dataset* pasien, iterasi berhenti pada iterasi ke 2 karena nilai pada masing masing *cluster* sama dengan iterasi ke 1 sehingga tidak diperlukan lagi perhitungan jarak dan menandakan bahwa data pasien pada tiap anggota *cluster* telah optimal dan dapat dilihat pada Tabel 7.

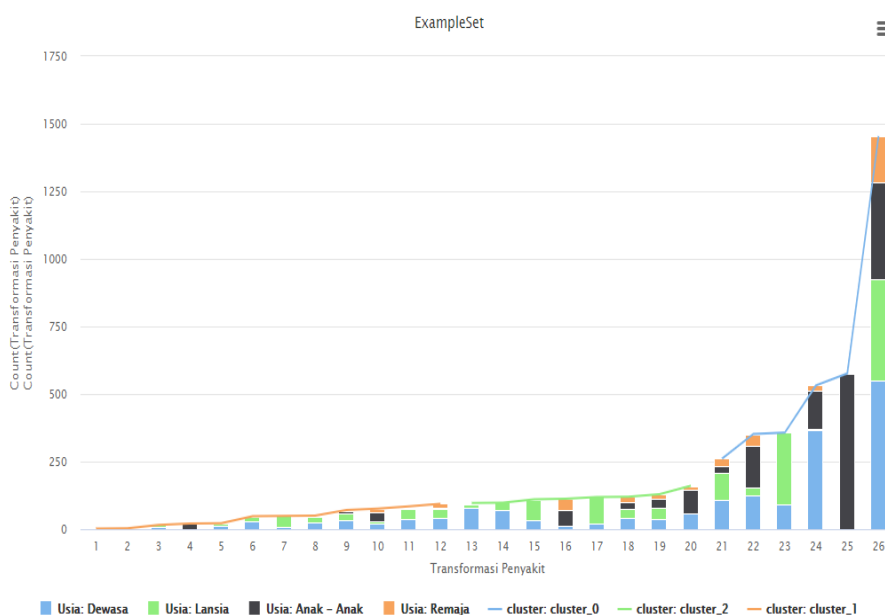
Tabel 7 Hasil perhitungan sampel data pasien pada iterasi 2

Pasien	Jarak <i>cluster</i> 1	Jarak <i>cluster</i> 2	Jarak <i>cluster</i> 3	Hasil <i>Cluster</i>
1	0,72111	13,0522	17,02116	1
2	0,848528	13,04454	17,01529	1
3	0,848528	13,04454	17,01529	1
4	0,848528	13,04454	17,01529	1
5	0,72111	13,00615	17,00941	1
183	0,565685	13,01384	17,01529	1
184	0,565685	13,01384	17,01529	1
185	0,565685	13,01384	17,01529	1
186	0,565685	13,01384	17,01529	1
187	0,565685	13,01384	17,01529	1
2736	25,0104	12,00666	8,019975	3
2737	24,01083	11,00727	7,02282	3
2851	23,00696	10,01798	6,043178	3
3211	13,01998	0,4	4,039802	2
3212	13,01998	0,4	4,039802	2
3213	13,01998	0,4	4,039802	2
3214	13,0123	0,6	4,06448	2
3215	13,0123	0,6	4,06448	2
3332	23,0113	10,06777	6,059703	3
4031	23,0113	10,06777	6,059703	3

3.2 Implementasi *Rapidminer*

Dalam implementasi algoritma K-Means clustering pada aplikasi rapidminer dataset kunjungan pasien pada tahun 2020 dan 2021 yang telah melalui tahapan data mining kemudian dilakukan clustering pada aplikasi rapidminer sehingga didapatkan hasil sebagai berikut :

1. *Dataset* kunjungan pasien tahun 2020



Gambar 4. Visualisasi *dataset* kunjungan pasien tahun 2020

Terdapat 5015 pasien dan 26 jenis penyakit pada tahun 2020 dimana pada cluster 0 yang merupakan *cluster* dengan penderita tertinggi terdapat 6 jenis penyakit yaitu transformasi penyakit 21-26 dengan 2 rentang usia yang paling tinggi yaitu usia anak – anak dengan penderita sebanyak 1259 dan dewasa sebanyak 1246 kemudian disusul dengan usia lansia sebanyak 770 dan terakhir remaja sebanyak 258. Lalu pada *cluster* 2 yang merupakan *cluster* dengan penderita sedang, terdapat 7 jenis penyakit dengan transformasi penyakit 13-20 dengan rentang usia yang paling tinggi yaitu dewasa dengan penderita sebanyak 357, lansia sebanyak 285, anak – anak sebanyak 203 dan remaja sebanyak 101 dan terakhir *cluster* 1 terdapat 12 jenis penyakit dengan penderita sedikit dengan 12 jenis juga didapat 2 rentang usia tertinggi yaitu dewasa sebanyak 225 pasien,

lansia sebanyak 200, anak – anak sebanyak 71 dan remaja sebanyak 40 pasien, sehingga didapatkan hasil 6 penyakit dan rentang usia terbanyak pada tahun 2020 yang dapat dilihat pada tabel 8.

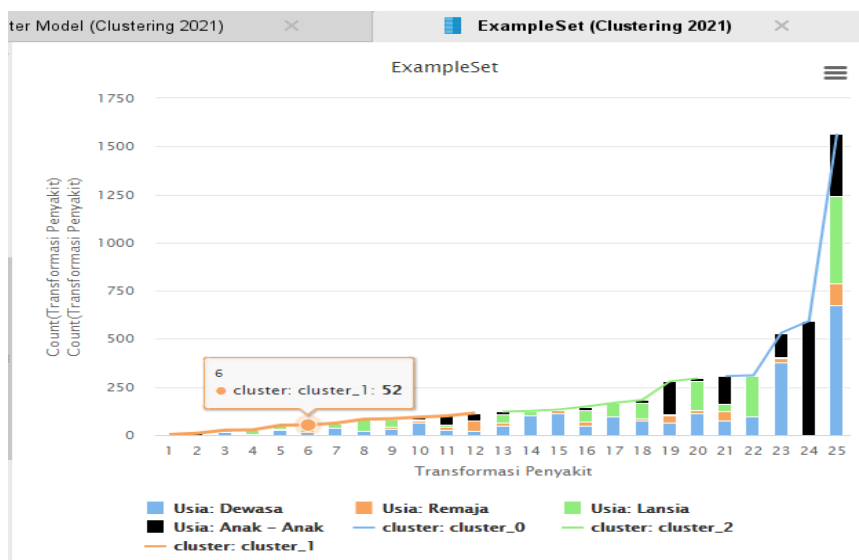
Tabel 8 Jenis penyakit dan rentang usia pasien terbanyak pada tahun 2020

No	Jenis Penyakit	Rentang Usia	Jumlah Penderita
1	Asma	Anak - Anak	25
		Remaja	27
		Dewasa	108
		Lansia	101
2	Diare	Anak - Anak	157
		Remaja	42
		Dewasa	126
		Lansia	27
3	Penyakit tekanan darah tinggi	Anak - Anak	0
		Remaja	0
		Dewasa	93
		Lansia	264
4	Tuberkulosis	Anak - Anak	142
		Remaja	19
		Dewasa	368
		Lansia	0
5	Imunisasi	Anak - Anak	576
		Remaja	0
		Dewasa	0
		Lansia	0
6	Demam Batuk Pilek	Anak - Anak	359
		Remaja	170
		Dewasa	551
		Lansia	375

2. Dataset kunjungan pasien tahun 2021

Pada tahun 2021 terdapat 5466 data pasien dengan 25 jenis penyakit, transformasi jenis penyakit 21-25 berada pada *cluster* 0 yang merupakan *cluster* dengan penderita tertinggi didapatkan rentang usia dewasa sebanyak 1229 pasien, anak-anak sebanyak 1188 pasien, lansia sebanyak 707 pasien dan remaja sebanyak 183 pasien, lalu pada *cluster* 2 didapatkan transformasi penyakit 13-20 yang merupakan *cluster* dengan penderita sedang dimana rentang usia dewasa menjadi rentang usia paling tinggi pada *cluster* 2 sebanyak 662 pasien, lansia sebanyak 451 pasien, anak – anak sebanyak 230 pasien dan terakhir pasien remaja sebanyak 109. Dan pada *cluster* 1 didapatkan transformasi jenis

penyakit 1-12 yang merupakan *cluster* dengan penderita sedikit memiliki pasien dewasa sebanyak 272, pasien lansia sebanyak 228, pasien anak – anak sebanyak 114 dan pasien remaja sebanyak 93 dimana didapatkan hasil 6 penyakit dan rentang usia terbanyak pada tahun 2021 yang dapat dilihat pada tabel 9.



Gambar 4. Visualisasi *dataset* kunjungan pasien tahun 2021

Tabel 9 Jenis penyakit dan rentang usia pasien terbanyak pada tahun 2021

No	Jenis Penyakit	Rentang Usia	Jumlah Penderita
1	Asma	Anak - Anak	13
		Remaja	14
		Dewasa	114
		Lansia	154
2	Diare	Anak - Anak	144
		Remaja	48
		Dewasa	78
		Lansia	36
		Anak - Anak	0
		Remaja	0
3	Penyakit tekanan darah tinggi	Dewasa	97
		Lansia	213
4	Tuberkulosis	Anak - Anak	128

		Remaja	19
		Dewasa	380
		Lansia	4
		Anak - Anak	594
5	Imunisasi	Remaja	0
		Dewasa	0
		Lansia	0
		Anak - Anak	322
6	Demam Batuk Pilek	Remaja	116
		Dewasa	674
		Lansia	454

Pada setiap tahunnya penyakit tertinggi pada tahun 2020 dan 2021 adalah demam, batuk, pilek, imunisasi, tuberkulosis, penyakit tekanan darah tinggi, diare dan asma. Sedangkan keguguran, Pendarahan pada kehamilan, persalinan dan nifas, Gizi Buruk, Kelainan Kornea dan refraksi merupakan penyakit yang paling sedikit diderita, Rentang usia dewasa merupakan penderita terbanyak pada setiap tahunnya kemudian disusul dengan anak-anak dan lansia, kemudian remaja merupakan rentang usia pasien yang paling sedikit melakukan pemeriksaan pada Puskesmas Muara Enim.

4. KESIMPULAN

Berdasarkan hasil dari visualisasi *dataset* kunjungan pasien pada tahun 2020 dan 2021 pada Puskesmas Muara Enim dengan metode algoritma K-Means pada aplikasi *rapidminer*, setelah dilakukan pemodelan data kemudian dianalisa, maka hasil pengujian pada penelitian ini menemukan bahwa :

- a) Dalam penerapan *dataset* kunjungan pasien dengan memanfaatkan algoritma *K-Means* dan dibantu aplikasi *rapidminer*, data pasien dapat dibagi kedalam 3 *cluster* dimana *cluster 0* merupakan *cluster* dengan penyakit dan penderita tertinggi, *cluster 2* merupakan *cluster* dengan penyakit dan penderita sedang sementara pada *cluster 1* merupakan *cluster* dengan penyakit dan penderita sedikit pada Puskesmas Muara Enim dari tahun 2020 hingga 2021.
- b) Berdasarkan dari hasil *cluster* pada *dataset* kunjungan pasien, dapat disimpulkan bahwa kelompok rentang usia tertinggi mengidap penyakit pada Puskesmas Muara Enim adalah Pasien Dewasa dan Anak – Anak kemudian disusul dengan pasien lansia dan terakhir pasien remaja pada setiap tahunnya.

- c) Berdasarkan *cluster* 0 yang merupakan *cluster* dengan penyakit dan penderita tertinggi, penyakit Asma, Diare, Penyakit tekanan darah tinggi, Tuberkulosis, Imunisasi dan Demam, batuk pilek merupakan penyakit yang memiliki penderita tertinggi pada Puskesmas Muara Enim dari tahun 2020 dan 2021.
- d) Dari penelitian ini hasil *clustering dataset* kunjungan pasien pada tahun 2020 dan 2021 diharapkan dapat menjadi acuan bagi Dinas Kesehatan dan Puskesmas Muara Enim dalam melakukan kegiatan penyuluhan bagi masyarakat khususnya pada kota Muara Enim, dan juga dalam pendistribusian stok obat ke Puskesmas Muara Enim, berdasarkan penyakit dengan penderita tertinggi, sedang dan sedikit.

DAFTAR PUSTAKA

- [1] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma k-means untuk pengelompokan penyakit pasien pada puskesmas cigugur tengah," *Journal of Information Technology*, pp. 39-44, 2020.
- [2] Dr. J. Leimena, "Peletak Konsep Dasar Pelayanan Kesehatan Primer (Puskesmas)" Depkes, 2005.
- [3] T. Tanty, B. S. Ginting, and M. Simanjuntak. "Pengelompokan Penyakit Pada Pasien Berdasarkan Usia Dengan Metode K-Means Clustering (Studi Kasus: Puskesmas Bahorok," *ALGORITMA: JURNAL ILMU KOMPUTER DAN INFORMATIKA*, 2021.
- [4] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi: Jurnal Sistem Informasi*, pp. 238-249, 2018.
- [5] A. Ali, "Klusterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo." *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, pp. 186-195, 2019.
- [6] P. Silitonga, I. S. Morina, "Klusterisasi pola penyebaran penyakit pasien berdasarkan usia pasien dengan menggunakan K-Means clustering," *Jurnal TIMES*, pp. 22-25, 2017.
- [7] N. Purba, Ponigsih, H. S. Tambunan, "Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau," *Journal of Information System Research (JOSH)*, pp. 220-226, 2021.
- [8] N. Y. Wardani, N. N. Murni, S. S. P. Luka, and G. Indrawan, "Analisis Penerapan K-means Untuk Pengelompokan Diagnosa Penyakit Kulit dan Kelamin Berdasarkan Rentang Usia," *SENAPATI 2016*, 2016.

- [9] A. Ramadhan, "Kategori Umur Menurut Depkes RI Tahun 2009," Departemen Kesehatan RI, 2014.
- [10] Hukum, Biro, and B. P. K. P. Humas, "Peraturan Pemerintah Nomor 43 Tahun 2004 Tentang Pelaksanaan Upaya Peningkatan Kesejahteraan Sosial Lanjut Usia," 2004.
- [11] A. Muzakir, H. Syaputra, and F. Panjaitan, "A Comparative Analysis of Classification Algorithms for Cyberbullying Crime Detection: An Experimental Study of Twitter Social Media in Indonesia," *Sci. J. Informatics; Vol 9, No 2 Novemb. 2022* DOI doi:10.15294/sji.v9i2.35149